Geometry and Topology in Electronic Structure Theory

Raffaele Resta
http://www-dft.ts.infn.it/~resta/

Notes subject to ongoing editing
This version run through LaTeX on 22–Oct–12 at 10:31

Contents

Chapter 1

Introduction

1.1 Why writing these Notes?

The occasion, or I should better say the pretext, for writing these Notes is a graduate course delivered at SISSA (International School for Advanced Studies, Trieste) in 2012 and covering a part—and a part only—of the topics dealt with here.

The need for some support material to this course is hardly compelling, since many very good review papers exist, whose union covers about the whole of the present topics, all of them state-of-the-art at the time of their publication. Some of these reviews are authored or coauthored by me [1, 2, 3, 4, 5, 6, 7, 8, 9], and many more by outstanding colleagues. Of these, I quote only the most recent ones; a non exhaustive list is [10, 11, 12, 13, 14, 15, 16, 17, 18].

The writing of a set of Lecture Notes in due form appears as a heavy burden, diverting for a significant amount of time the author from his day-to-day research work, already also diverted by administration and teaching. Instead, I used above the word "pretext", because these notes are mostly written for my own sake. In fact I regard such burden as an important occasion for a pause of reflection, devoted to critical rethinking about the subject in its wholeness. In the present case, the pause of reflection was made possible by an extended stay (four months) at CECAM in Lausanne.

I wrote Lecture Notes in four other occasions, last time in 2000 [19]. I know by experience that the task of writing in uniform notations and in a logical sequence many results scattered in the literature leads me to scrutinize from a novel viewpoint not only other people's work, but even my own. Here LaTeX plays a major role; soon after its appearance in the early 1980s my way of thinking has changed dramatically. While previously I—like everybody—reasoned by jotting formulas on a sheet of paper, nowadays I cannot reason clearly unless the formulas are neatly typeset in LaTeX.

The very fact of writing many different known results altogether has the beneficial effect of making quite evident several links, hitherto unsuspected. The perspective view gained while writing the present Notes (and typesetting them in LaTeX) will doubtless influence the future course of my research activity, as in fact already happened in the previous occasions.

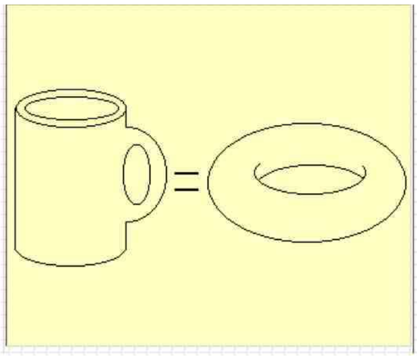

Figure 1.1: The hallmark of topology, as in many popular presentations. Hundreds of figures like this, and even some very perspicuous videos, can be downloaded from the internet. The two closed surfaces ("two-dimensional compact manifolds") have the same topological invariant $g = 1$, which measures the number of handles.

1.2 What topology is about

Topology is defined as a branch of mathematics that describes properties which remain unchanged under smooth deformations; such properties are usually labelled by integer numbers, named topological invariants. The concepts and tools belonging to topology are continuity and connectivity, open and closed sets, neighborhoods, and the like.

Differentiability, or even a metric structure, are not needed; theorems are proved under very general hypotheses, and are therefore very powerful, being applicable to very diverse frameworks. The tradeoff is that proofs, and even definitions, look clumsy and obscure to readers with the mathematical background of a typical condensed matter physicist. The good news is that the topological properties most relevant for electronic structure theory can be formulated in the more familiar language of differential geometry.

Many introductions to topology start with the statement that, to a topologist, a coffee cup and a doughnut are the same thing, as in Fig. 1.1. Intuitively, the common feature of the two objects is the presence of one, and only one, handle. The mathematical definition of "handle" is coming soon.

1.2.1 Gauss-Bonnet theorem

We start with the simplest example, a sphere, and a tangent plane at a given point. In a local system of Cartesian coordinates on the plane the equation of the sphere is

$$z = R - \sqrt{R^2 - x^2 - y^2} \simeq \frac{x^2 + y^2}{2R}, \tag{1.1}$$

and the Hessian matrix is

$$H = \begin{pmatrix} \frac{\partial^2 z}{\partial x^2} & \frac{\partial^2 z}{\partial x \partial y} \\ \frac{\partial^2 z}{\partial y \partial x} & \frac{\partial^2 z}{\partial y^2} \end{pmatrix} = \begin{pmatrix} 1/R & 0 \\ 0 & 1/R \end{pmatrix}. \tag{1.2}$$

The Gaussian curvature $\mathbf{\Omega}$ is by definition the determinant of the Hessian at the tangency point. It is obviously constant and equal to $1/R^2$ at any point of the sphere; notice that the orientation of the z axis (either inwards or outwards) is irrelevant. The integral of $\mathbf{\Omega}$ over the whole closed surface is 4π.

Next we consider a smooth (i.e. twice differentiable) closed surface of arbitrary shape: the Gaussian curvature is defined as the determinant of the Hessian

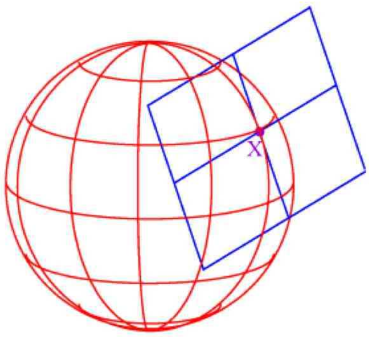

Figure 1.2: A sphere of radius R, and its tangent plane in a generic point. The Gaussian curvature in this trivial case is $\mathbf{\Omega} = 1/R^2$.

at the tangent plane, similarly to what we did for the sphere:

$$\mathbf{\Omega} = \det \begin{pmatrix} \frac{\partial^2 z}{\partial x^2} & \frac{\partial^2 z}{\partial x \partial y} \\ \frac{\partial^2 z}{\partial y \partial x} & \frac{\partial^2 z}{\partial y^2} \end{pmatrix} . \tag{1.3}$$

In general, $\mathbf{\Omega}$ can be positive, negative (at a saddle point), or zero (e.g. for a plane or a cylinder). The Gauss-Bonnet theorem states that for any closed smooth surface

$$\frac{1}{2\pi} \int_S d\sigma \, \mathbf{\Omega} = 2(1-g), \tag{1.4}$$

where g is a nonnegative integer, called the "genus" of the surface. Surfaces which can be continuously deformed into each other (i.e. "homeomorphic") have the same genus. For the sphere and any surface homeomorphic to it $g = 0$; both the coffee cup and the doughnut, Fig. 1.1 have $g = 1$; a double-handle cup has $g = 2$. The genus is thus the mathematical definition for the number of handles.

1.2.2 Euler characteristic

We have considered smooth surfaces so far, but topological invariants are based on the more general condition of continuity, and—to the delight of mathematicians—unsurprisingly can be defined even for pathological surfaces ("manifolds" in topology-speak). The simplest non smooth case addresses polyhedra, where the Gaussian curvature is either zero (on the faces) or singular (at vertices and edges).

The Euler characteristic is defined as $\chi = V - E + F$, where V is the number of vertices, E is the number of edges, and F is the number of faces. If we address the set of regular polyhedra (tetrahedron, cube, octahedron, dodecahedron, icosahedron) it is easily verified that $\chi = 2$. All these surfaces can be

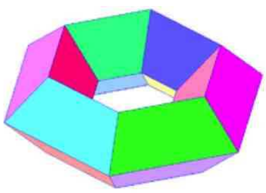

Figure 1.3: A doughnut shaped polyhedron. This surface has Euler characteristic $\chi = 0$ or, equivalently, genus $g = 1$.

continuously deformed into ("are homeomeorphic to") each other, and into a sphere. In fact there is a one-to-one relationship between the Euler characteristic and the genus: $\chi = 2(1 - g)$. Polyhedra can also have $\chi \neq 2$, like the doughnut-shaped one shown in Fig. 1.3.

1.3 Electronic wavefunctions

In the domain of electronic structure, the typical object addressed via geometrical and/or topological concepts is the electronic ground state of some system. Whenever an observable effect has the nature of a topological invariant, i.e. it is an integer number, two remarkable features occur. (1) The observable is measurable in principle with infinite precision (10^{-9} is actually attained for the quantum Hall effect). (2) The observable is very robust under even strong variations of the sample conditions; a very disruptive perturbation is needed to switch from one integer to another. Topology concerns mostly insulators: in this case the disruptive perturbation amounts to crossing a metallic state.

These Notes are entirely devoted to physical properties having a topological and/or geometrical character. I am not sure of always using the right semantics. Loosely speaking, I would use the term "topological" for something which is quantized, and "geometrical" for something which is not. The framework and the mathematical tools are often the same for quantized and nonquantized quantities, the former frequently occurring as special cases of the latter.

The Berry phase is the typical geometrical quantity which is not quantized, although it can be quantized in high-symmetry cases. The macroscopic polarization of a solid is a Berry phase, and is obviously (from an experimental viewpoint) a nonquantized observable. Nonetheless, there are aspects of the modern theory of polarization that I would define topological. The same applies to other geometrical properties considered in this Notes; it is reassuring that even other authors often use "geometrical" and "topological" as synonymous.

Finally, a few words about the many calculations cited and sometimes briefly outlined here. Unless otherwise stated, the term "first-principle calculations", when referred to a condensed matter system, means density functional calculations; independent-electron eigenfunctions and eigenvalues are the Kohn-Sham (KS) ones. Despite these Notes mostly address a computational physics readership, no technical details are given (basis sets, pseudopotentials, functionals...); they are obviously detailed in the original literature, while the focus here is on the physical properties.

1.4 Units

We use Gaussian electromagnetic units throughout: these have the advantage (at variance with SI units) that electric and magnetic fields have the same dimensions. Furthermore, the nasty ε_0 and μ_0 disappear; SI formulas are converted by setting $4\pi\varepsilon_0 = 1$ and $4\pi/\mu_0 = 1$.

For a single particle, the Newton equation of motion and the Hamiltonian read, respectively

$$M\frac{d\mathbf{v}}{dt} = \mathbf{f} = Q\left(\mathbf{E} + \frac{1}{c}\mathbf{v} \times \mathbf{B}\right), \tag{1.5}$$

$$H = \frac{1}{2M}\left(\mathbf{p} - \frac{Q}{c}\mathbf{A}(\mathbf{r})\right)^2 + Q\Phi(\mathbf{r}). \tag{1.6}$$

Generally, Gaussian electromagnetic units are associated to mechanical cgs units, but this is no means necessary. In electronic structure theory, it is expedient to associate Gaussian electromagnetic units with atomic units (a.u.), defined as $e^2 = 1$, $m_\mathrm{e} = 1$, $\hbar = 1$. The unit of energy is the hartree (1 Ha = 2 Ry = 27.21 eV). In the present Notes the electron charge is $-e$, with $e > 0$; this sign choice agrees with most (but not all) the recent literature. For instance, the very popular review of Ref [1] adopts $e < 0$.

The speed of light in a.u. is $c = 137$. This immediately hints at why the largest atomic number Z in the periodic table is $Z \simeq 100$: in fact the core electrons have (in a.u.) energies of the order of Z^2, hence velocities of the order of Z.

1.5 Symbols

I am faced here with two contrasting issues: adopting the symbols most currently used in the literature, and adopting different symbols for different objects. This proved to be near to impossible in a work of the present kind, if baroque symbols are to be ruled out. For instance along the present Notes I do use $A, \mathbf{A}, \mathcal{A}, \boldsymbol{\mathcal{A}}, \mathscr{A}$, all with a different meaning. Similarly, I use $P, \hat{P}, \mathbf{P}, \mathcal{P}, \mathscr{P}$. Despite this, I found unavoidable to use—in different Chapters—the same symbol for different objects. For instance, depending on the context, the symbol P may indicate a projector or, otherwise, a one-dimensional electrical polarization. Another example is the symbol $\boldsymbol{\Omega}$ used for the Berry curvature, while Ω_I is the gauge-invariant quadratic spread of the Wannier functions. Therefore caution is in order when extrapolating a given symbol from its own context.

1.6 Gauge and flux

We consider here a simple exercise which plays the role of a very important paradigm; it illustrates basic concepts and results which are going to reappear several times all along the present Notes.

We address the single-particle Hamiltonian

$$H = \frac{1}{2m}(\mathbf{p} + \frac{e}{c}\mathbf{A})^2 + V(\mathbf{r}), \tag{1.7}$$

where the vector potential $\mathbf{A}$ is independent of space and time. It is usually said that $\mathbf{A}$ is a pure gauge, meaning with this that it does not affect the fields:

$$\mathbf{B} = \nabla \times \mathbf{A} \equiv 0, \qquad \mathbf{E} = -\frac{1}{c}\frac{\partial \mathbf{A}}{\partial t} \equiv 0. \tag{1.8}$$

1.6.1 Classical mechanics

Let us first adopt a classical viewpoint. The Hamilton equation of motions are

$$\dot{\mathbf{p}} = -\frac{\partial H}{\partial \mathbf{r}} = -\nabla V(\mathbf{r}) \tag{1.9}$$

$$\dot{\mathbf{r}} = \frac{\partial H}{\partial \mathbf{p}} = \frac{1}{m}(\mathbf{p} + \frac{e}{c}\mathbf{A}). \tag{1.10}$$

From these we get

$$\mathbf{p} = m\dot{\mathbf{r}} - \frac{e}{c}\mathbf{A}, \tag{1.11}$$

which leads to the Newton equation of motion

$$m\ddot{\mathbf{r}} = -\nabla V(\mathbf{r}). \tag{1.12}$$

The bottom line looks quite obvious: a pure gauge has no effect. A basic tenet of classical mechanics is that the equations of motion can always be directly expressed in terms of the forces (i.e. the fields), while the potentials—scalar and vector—are auxiliary quantities, devoid of physical meaning.

1.6.2 Quantum mechanics, open boundary conditions

Next we switch to quantum mechanics. It is expedient to rewrite Eq. (1.7) as

$$H(\boldsymbol{\kappa}) = \frac{1}{2m}(\mathbf{p} + \hbar\boldsymbol{\kappa})^2 + V(\mathbf{r}), \qquad \boldsymbol{\kappa} = \frac{e}{c\hbar}\mathbf{A}, \tag{1.13}$$

where $\boldsymbol{\kappa}$, having the dimensions of an inverse length, will be referred to as "twist" in the following. The Schrödinger equation is

$$H(\boldsymbol{\kappa})|\psi_n(\boldsymbol{\kappa})\rangle = \epsilon_n(\boldsymbol{\kappa})|\psi_n(\boldsymbol{\kappa})\rangle. \tag{1.14}$$

The eigenvectors and eigenvalues of the Schrödinger equation depend on the boundary conditions assumed.

The so-called open boundary conditions (OBCs) require that the bound eigenstates are square-integrable over $\mathbb{R}^3$. Let $|\psi_n(0)\rangle$ be a nondegenerate eigenstate of the "untwisted" Hamiltonian within OBCs. Then the state $\mathrm{e}^{-i\boldsymbol{\kappa}\cdot\mathbf{r}}|\psi_n(0)\rangle$ obviously obeys OBCs as well, and also obeys Eq. (1.14) with a $\boldsymbol{\kappa}$-independent eigenvalue. Therefore it coincides with the n-th eigenstate $|\psi_n(\boldsymbol{\kappa})\rangle$ of the twisted Hamiltonian; notice that this eigenstate is arbitrary by a $\boldsymbol{\kappa}$-dependent phase factor.

We conclude that a pure gauge within OBCs affects the wavefunction, but does not affect any of the observable quantities, such as expectation values, density, and current. We spell this, in jargon, by saying that the "twist" is easily "gauged away" within OBCs.

1.6.3 Quantum mechanics, periodic boundary conditions

We assume periodic boundary conditions (PBCs) over a cubic box of side L, i.e. we require the eigenstates of Eq. (1.14) to be Born-von-Kàrmàn periodic with period L over x, y, and z at any given $\boldsymbol{\kappa}$. Each Cartesian coordinate is therefore equivalent to an angle, e.g. $\varphi_x = 2\pi x/L$.

If $|\psi_n(0)\rangle$ is an eigenstate of the untwisted Hamiltonian within PBCs, then the state $\mathrm{e}^{-i\boldsymbol{\kappa}\cdot\mathbf{r}}|\psi_n(0)\rangle$ obeys Eq. (1.14) with a $\boldsymbol{\kappa}$-independent eigenvalue, but for a general $\boldsymbol{\kappa}$ it *does not* obey PBCs, and therefore in general does not coincide with the genuine eigenstate $|\psi_n(\boldsymbol{\kappa})\rangle$. Within PBCs the spectrum of Eq. (1.14) depends on the twist $\boldsymbol{\kappa}$ in a nontrivial way.

If $|\psi_n(\boldsymbol{\kappa})\rangle$ is an eigenstate of Eq. (1.14) within PBCs with eigenvalue $\epsilon_n(\boldsymbol{\kappa})$, then the auxiliary state $|\tilde{\psi}_n(\boldsymbol{\kappa})\rangle = \mathrm{e}^{i\boldsymbol{\kappa}\cdot\mathbf{r}}|\psi_n(\boldsymbol{\kappa})\rangle$ obeys the untwisted ($\boldsymbol{\kappa} = 0$) Schrödinger equation, and quasi-periodical (a.k.a. "twisted" or "skewed")

boundary conditions: at any two opposite faces of the cube the wavefunction differs by a $\boldsymbol{\kappa}$-dependent phase factor.

In other words the problem can be formulated in two equivalent ways: either the Hamiltonian is $\boldsymbol{\kappa}$-dependent, as in Eq. (1.14), and the boundary conditions are $\boldsymbol{\kappa}$-independent; or the Hamiltonian is $\boldsymbol{\kappa}$-independent but the boundary conditions are "twisted" in a $\boldsymbol{\kappa}$-dependent way.

1.6.4 Example: Free particle in 1d

For the sake of simplicity, we consider Eq. (1.14) in 1d, and with $V \equiv 0$:

$$\frac{\hbar^2}{2m}\left(-i\frac{d}{dx}+\kappa\right)^2 |\psi_n(\kappa)\rangle = \epsilon_n(\kappa)|\psi_n(\kappa)\rangle. \tag{1.15}$$

The eigenfunctions within PBCs and the spectrum are

$$\langle x|\psi_n(\kappa)\rangle \propto \mathrm{e}^{i\frac{2\pi n}{L}x}, \qquad n \in \mathbb{Z}, \tag{1.16}$$

$$\epsilon_n(\kappa) = \frac{\hbar^2}{2m}\left(\frac{2\pi n}{L}+\kappa\right)^2, \tag{1.17}$$

where the nontrivial κ-dependence is perspicuous. The velocity operator can be written as

$$v = \frac{1}{\hbar}\frac{\partial H}{\partial \kappa}, \tag{1.18}$$

and the Hellmann-Feynman theorem yields

$$\langle\psi_n|v|\psi_n\rangle = \frac{1}{\hbar}\frac{d\epsilon_n(\kappa)}{d\kappa}. \tag{1.19}$$

We have introduced PBCs as a basic framework of condensed matter physics. Many concepts (like the Bloch vector or the Fermi surface) make sense only within PBCs. But we also may regard this problem as if the electrons were confined to a circular rail of circumference L, as in Fig. 1.4. There is no field (electric or magnetic) on the rail, but a constant vector potential of intensity $A = c\hbar\kappa/e$ is present along the rail; eigenvectors and eigenfunctions *depend* on its value.

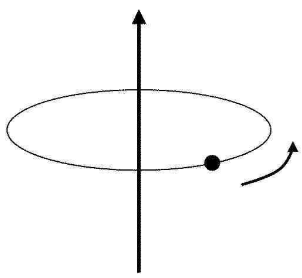

Figure 1.4: The electron motion is confined to a circular rail. A constant vector potential $A = c\hbar\kappa/e$ along the rail, as in Eq. (1.15), corresponds to vanishing fields (electric and magnetic), yet the spectrum *depends* on the "inaccessible flux" threading the surface encircled by the rail.

1.6.5 Flux and flux quantum

The constant vector potential A on the circular rail corresponds to a magnetic flux $\phi = LA$ threading the surface encircled by the rail, in a region *not visited* by the electronic system; it has been appropriately called by some authors "inaccessible flux".

We further observe that the spectrum, Eq. (1.17), is periodic in κ with period $2\pi/L$; alternatively, it is periodic in the flux ϕ with period $\phi_0 = 2\pi\hbar c/e = hc/e$, the elementary flux quantum. In cgs units $hc/e = 4.135 \times 10^{-7}$ gauss cm^2, while in SI units $\phi_0 = h/e = 4.136 \times 10^{-15}$ Wb. Notice also that, in the framework of superconductivity, the same symbol ϕ_0 indicates *one half* of this (it refers to electron pairs).

We stress that only the fractional part of the flux affects the results in a nontrivial way. This is perspicuous if we recast Eq. (1.17) as

$$\epsilon_n(\phi) = \frac{\hbar^2}{2m}\left(\frac{2\pi}{L}\right)^2\left(n + \frac{\phi}{\phi_0}\right)^2. \tag{1.20}$$

The flux breaks time-reversal symmetry ($\kappa \to -\kappa$), and the spectrum is nondegenerate, except when $\phi = 0$ or $\phi = \phi_0/2$, the latter also called "π flux". In these two cases (and in these cases only) the eigenfunctions can be chosen as real.

When the flux is varied with time, an emf is induced along the loop. Using Eq. (1.19), the current is

$$I = -ev = -c\frac{d\epsilon_n}{d\phi}. \tag{1.21}$$

This result is remarkable: it holds even in presence of a potential $V(x)$, and generalizes straightforwardly to N noninteracting electrons. It will be used in the discussion of the quantum Hall effect: see Eq. (2.17) below.

Chapter 2

Early discoveries

2.1 The Aharonov-Bohm effect: A paradox?

The Aharonov-Bohm effect is the paradigm for a measurable effect induced by an inaccessible flux. We anticipate that in many other phenomena such flux may be purely "geometrical" or "topological", without any relationship to a genuine magnetic field: this is e.g., the case considered in the next Section. It is only in the Aharonov-Bohm effect that one addresses indeed the inaccessible flux of a magnetic field, as present e.g. inside a solenoid. An interference experiment detects the presence of the flux even when the electronic motion is confined in the region outside the solenoid, where the magnetic field is zero. This seems paradoxical: something which "happens" in a region not visited by the quantum particle may affect some observable properties. Indeed, the founding fathers of quantum mechanics (in the 1920s) failed to notice such peculiar feature. It only surfaced more than 30 years afterwards in the milestone paper by Aharonov and Bohm [20], appeared in 1959, whose abstract states verbatim *"...contrary to the conclusions of classical mechanics, there exist effects of potential on charged particles, even in the regions where all the fields (and therefore the forces on the particles) vanish"*.

The paper was shocking, and its conclusions were challenged by several authors; nonetheless experimental validations appeared as early as 1960 [22, 23]. The main message of Ref. [20] is at the basis of many subsequent developments

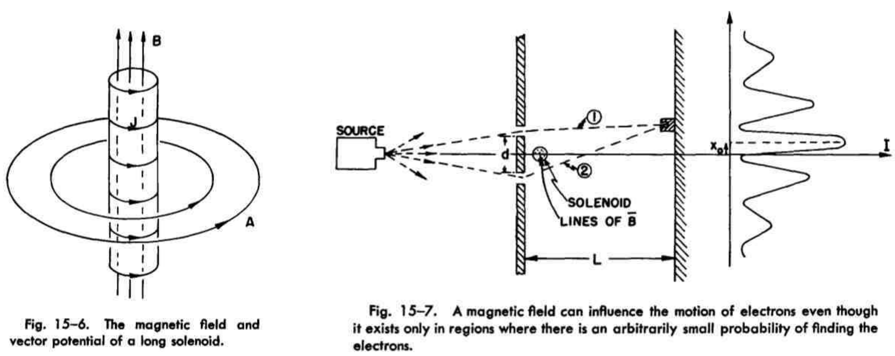

Figure 2.1: The Aharonov-Bohm interference experiment (From Ref. [21])

in electronic structure theory, many of them illustrated below in the present Notes. The Aharonov-Bohm effect is also at the root of the commercial SQUID technology [24].

It is remarkable that R. P. Feynman included the Aharonov-Bohm effect in his legendary lectures, delivered to the sophomore class at Caltech during the 1962-63 academic year [21]. In the final sentence about this topic, Feynman says: *"...$\mathbf{E}$ and $\mathbf{B}$ are slowly disappearing from the modern expression of physical laws; they are being replaced by $\mathbf{A}$ and Φ"*.

It is also remarkable and shameful that a paper bearing the title "Nonexistence of the Aharonov-Bohm effect" [25] was published as late as 1978. All challenges disappeared with the publication in 1984 of the celebrated paper by Michael Berry (now *Sir* Michael Berry [26]), where the eponymous phase made its first appearance [27].

2.2 Conical intersections in molecules

At the time Berry wrote his famous paper, only two occurrences of a geometrical phase (called Berry phase soon afterwards) in quantum mechanics were known to him: the Aharonov-Bohm effect and a somewhat exotic phenomenon occurring in molecular physics. Even the latter was known since the late 1950s [28, 29], and appropriately rebaptized in the late 1970s as "molecular Aharonov-Bohm effect" [30, 31]. In the subsequent years Berry phases were discovered in many branches of physics.

The smallest molecular system where the molecular Aharonov-Bohm effect is possible is a trimer, having three internal coordinates (e.g. the three internuclear distances), and the simplest trimers are of course the homonuclear ones, where symmetry plays a major role. I give a simple outline for this particular system: a dynamical Jahn-Teller effect, bearing the conventional symmetry label $E \otimes \varepsilon$.

We focus on a trimer of monovalent atoms, e.g. H_3 or Na_3, and we assume an independent-electron picture in the Born-Oppenheimer approximation. We start with the molecule in the equilateral configurations, Fig. 2.2. Two of the valence electrons occupy a totalsymmetric orbital, while the unpaired electron occupies the next available one, which has E symmetry and is doubly degenerate. In a simple tight-binding (alias minimal-basis LCAO) scheme, a possible basis in the two-dimensional manifold is:

$$|1\rangle = \frac{1}{\sqrt{2}}(\ |\mathrm{B}\rangle - |\mathrm{C}\rangle\ )\ ; \qquad |2\rangle = \frac{1}{\sqrt{6}}(\ 2|\mathrm{A}\rangle - |\mathrm{B}\rangle - |\mathrm{C}\rangle\ )\ , \tag{2.1}$$

where A,B,C are atomic labels (as in the figure). This choice deserves an important comment. We are adopting OBCs, as appropriate for an isolated molecule, and the Hamiltonian is invariant under time reversal (no magnetic field, no spin-orbit interaction). These two conditions guarantee that the orbitals may always be chosen as real. They *may*, but they don't need: it may instead be convenient to choose a complex basis in the same two-dimensional degenerate manifold.

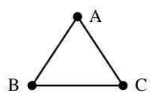

Figure 2.2: A homonuclear trimer in its equilateral configuration.

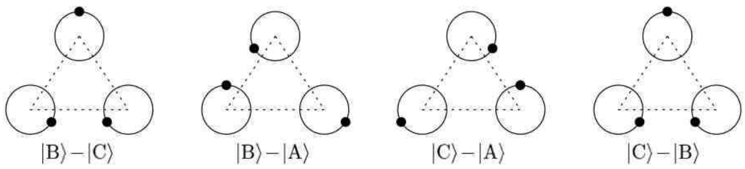

Figure 2.3: A schematic representation of a (counterclockwise) pseudorotation, where subsequent snapshots differ by $2\pi/3$. The corresponding electronic ground states in the tight-binding approximation, are also shown.

When we distort the molecule from its equilateral configuration, the doublet is linearly split: one of the two components is energetically favored, the molecule undergoes Jahn-Teller distortion, and the electronic ground state in the Born-Oppenheimer approximation becomes nondegenerate.

Next we analyze the motion of the nuclei. There are three linearly independent normal modes for the small oscillations of the internal coordinates. Of course, *in absence of a Jahn-Teller effect*, the equilateral configuration is the equilibrium one. One mode is totalsymmetric, and cannot split the electronic levels. The remaining modes are degenerate, having in fact E symmetry, and couple to the electronic doublet, originating in fact the dynamical Jahn-Teller effect. The notation $E \otimes \varepsilon$ means indeed that an E vibrational mode is coupled to an E electronic state: conventionally, one uses upper case letters as symmetry labels for the vibrational states, and lower case ones for the electronic states.

The adiabatic electronic ground state follows the nuclear motion. For a cyclic pseudorotation, shown in Fig. 2.3, the Hamiltonian is periodical, but the electronic wavefunction is *antiperiodical*. The total wavefunction in the Born-Oppenheimer approximation factors into the electronic one times the nuclear one. Given that the total wavefunction must be single-valued, even the nuclear wavefunction must be quantized using antiperiodical boundary conditions, and this affects the pseudorotation spectrum in a measurable way.

This feature has to do with the peculiar shape of the Born-Oppenheimer surface, shown in Fig. 2.4. If we adopt a two-dimensional Cartesian normal coordinate $\boldsymbol{\xi} = (\xi_1, \xi_2)$, the ionic displacements are parametrized as:

$$
\begin{aligned}
x_A &= \xi_1 & y_A &= \xi_2 \\
x_B &= -\tfrac{1}{2}\xi_1 + \tfrac{\sqrt{3}}{2}\xi_2 & y_B &= -\tfrac{\sqrt{3}}{2}\xi_1 - \tfrac{1}{2}\xi_2 \\
x_C &= -\tfrac{1}{2}\xi_1 - \tfrac{\sqrt{3}}{2}\xi_2 & y_C &= \tfrac{\sqrt{3}}{2}\xi_1 - \tfrac{1}{2}\xi_2
\end{aligned}
$$

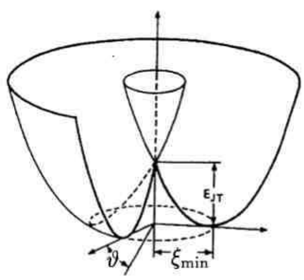

Figure 2.4: The Born-Oppenheimer surface of the Jahn-Teller split doublet: a double-valued function with a conical intersection. The potential minimum is a circle of radius $\xi_{\min}$ centered at the degeneracy point.

The meaning of this coordinate choice is transparent with reference to Fig 2.3: when the atom A is displaced by $\boldsymbol{\xi}$, the displacements of B and C are of equal magnitude $|\boldsymbol{\xi}|$, but pointing in directions rotated by $-2\pi/3$ and $-4\pi/3$, respectively. If we neglect Jahn-Teller coupling beyond linear order, no potential energy is associated to a motion at constant $|\boldsymbol{\xi}|$, which is indeed a free pseudorotation (or a "rotation wave"), also schematized in the succession of snapshots in Fig. 2.3.

In absence of Jahn-Teller coupling, the surface would simply be a parabola, everywhere doubly degenerate. The linear Jahn-Teller splitting is function of $|\boldsymbol{\xi}|$, hence to linear order the electronic eigenvalues are:

$$E_{\pm}(\boldsymbol{\xi}) \propto |\boldsymbol{\xi}|^2 \pm \text{const } |\boldsymbol{\xi}|. \tag{2.2}$$

This double-valued function is displayed in Fig. 2.4 and has a conical intersection at the origin. The lowest sheet $E_{-}(\boldsymbol{\xi})$ has a circular valley of radius $\xi_{\min}$, where a classical particle travels freely (if nonlinear Jahn-Teller coupling is neglected). Nothing exotic happens if the nuclear motion can be considered as classic; but when we quantize the nuclear degrees of freedom, *antiperiodical* boundary conditions have to be imposed for the cyclic motion, as said above.

A simple approximation for the rotovibrational levels is thus:

$$E(u,j) = (u + \frac{1}{2})\,\omega_0 + Aj^2, \tag{2.3}$$

corresponding to an oscillation of frequency ω_0 and quantum number u, and a two-dimensional internal rotation with rotor constant A. The antiperiodical boundary conditions imply half-odd-integer values for the quantum number j. The pseudorotation term in the spectrum can be compared to Eq. (1.20); the moment of inertia in the prefactor becomes obviously a nuclear rotor, but the spectrum is the same if we identify the inaccessible flux ϕ with *half* a flux quantum ϕ_0 (a.k.a. π flux).

There is no magnetic field in this problem; the flux is purely topological and can be regarded as an obstruction: the nuclear path cannot be contracted without crossing a degeneracy point. It is remarkable that the topological nature of this problem was clearly stated as early as 1963—much earlier than topology became fashionable in electronic structure—by Herzberg and Longuet-Higgins [29], who say verbatim: *"...a conically self-intersecting potential surface has a different topological character from a pair of distinct surfaces which happen to meet at a point. Indeed, if an electronic wave function changes sign when we move round a closed loop in configuration space, we can conclude that somewhere inside the loop there must be a singular point at which the wave function is degenerate"*.

In modern jargon, we would say that the cases $\phi = 0$ and $\phi = \phi_0/2$ are *topologically distinct*; owing to time-reversal invariance, other flux values are ruled out. The present paradigm also illustrates the robustness of topological properties against smooth deformations. For instance, here we have addressed the ultrasimple tight-binding model, but the ground wavefunction can be "continuously deformed" to the exact correlated wavefunction: topology-wise, the two wavefunctions are essentially the same object, insofar as the conical intersection is present. Notice also that at the conical intersection the Born-Oppenheimer approximation breaks down.

One could also address more general closed paths, according to their winding number round the obstruction. Only paths having the same winding number can be continuously deformed into each other: they are "homotopic".

2.3 Quantization of the surface charge

The pioneering selfconsistent calculations of the electronic structure of surfaces, performed at IBM (Yorktown Heights) and at Bell Labs in the mid 1970s, pointed out the occurrence of quantization of charge at insulating surfaces. After an early paper by V. Heine in 1966 [32], the theorem made its appearance in a 1974 paper by Appelbaum and Hamann [33]. Other papers addressed the issue in the 1970s [34, 35], but the topological explanation came much later; it will be discussed below, Sec. 5.3.4.

The quantization of surface charge may appear counterintuitive, if one sticks at the idea that a solid is an array of classical charges (ions), as many people still do. Possibly because of its counterintuitive content, this important theorem is surprisingly ignored even by well known specialists in surface physics. The extreme of such ignorance occurs in a recently published invited review paper—which I abstain from quoting—about polar surfaces. The theorem is even more ignored in quantum chemistry, where it addresses end charges in linear polymers.

Electrons are quantum particles, and classical ideas may prove wrong. Solids are *not* assemblies of ions; they are assemblies of atoms, having ionic character only because neighboring atoms have a different electronegativity [36]. At the surface, one has to look at what happens to the bonds.

A simple statement of the theorem is the following. If the bulk of the crystal is centrosymmetric, and if the surface is insulating, then the charge per surface cell may only be an integer or half integer; the surface charge can be nonquantized only if the bulk is noncentrosymmetric, or if the surface is metallic.

Quite often, the actual quantized value is zero because of energy considerations; therefore even polar surfaces are (counterintuitively) neutral under the above two essential hypotheses, which I stress again: the bulk is centrosymmetric, and the surface is insulating. The microscopic mechanism can be understood as an intrinsic surface-state neutralization [36]; however, topology guarantees quantization *independently* of microscopic details.

We nowadays regard bulk-surface correspondence in many phenomena as one of the hallmarks of geometry and topology in condensed matter physics. In modern jargon, I would say that the surface charge of insulators is "topologically protected".

2.4 Integer quantum Hall effect

2.4.1 Classical theory (Drude-Zener)

We consider any 2d system, in the setup shown in Fig. 2.5. If dissipation is accounted for by a single relaxation time τ, the Newton equation of motion for a single carrier of mass m and charge $-e$, is

$$m\left(\frac{d\mathbf{v}}{dt}+\frac{1}{\tau}\mathbf{v}\right)=-e\left(\mathbf{E}+\frac{1}{c}\mathbf{v}\times\mathbf{B}\right). \tag{2.4}$$

Setting $d\mathbf{v}/dt = 0$ we get the steady-state solution:

$$\mathbf{v} = -\frac{e\tau}{m}\left(\mathbf{E} + \frac{1}{c}\mathbf{v}\times\mathbf{B}\right). \tag{2.5}$$

In terms of the cyclotron frequency $\omega_c = \frac{eB}{mc}$ the solution with $E_y = 0$ is

$$\begin{aligned} v_x &= -\frac{e\tau}{m}E_x - \omega_c\tau v_x \\ v_y &= \omega_c\tau v_x. \end{aligned} \tag{2.6}$$

If n is the carrier density, the current is $\mathbf{j} = -ne\,\mathbf{v}$:

$$\begin{aligned} j_x &= \frac{ne^2\tau}{m}E_x - \omega_c\tau j_x \\ j_y &= \omega_c\tau j_x. \end{aligned} \tag{2.7}$$

in zero $\mathbf{B}$ field we retrieve the standard Drude (diagonal) conductivity:

$$j_x = \sigma_0 E_x, \qquad \sigma_0 = \frac{ne^2\tau}{m}, \tag{2.8}$$

while for $\mathbf{B} \neq 0$ the conductivity tensor is:

$$\begin{aligned} j_x &= \frac{\sigma_0}{1+(\omega_c\tau)^2}E_x = \sigma_{xx}E_x \\ j_y &= \frac{\omega_c\tau\sigma_0}{1+(\omega_c\tau)^2}E_x = \sigma_{yx}E_x. \end{aligned} \tag{2.9}$$

Inversion of the conductivity tensor

$$\rho_{xx} = \frac{\sigma_{xx}}{\sigma_{xx}^2 + \sigma_{yx}^2} \qquad \rho_{xy} = \frac{\sigma_{yx}}{\sigma_{xx}^2 + \sigma_{yx}^2} \tag{2.10}$$

provides a remarkably simple expression for the longitudinal and transverse resistivity

$$\rho_{xx} = 1/\sigma_0 = \frac{m}{ne^2\tau}, \quad \rho_{xy} = \frac{m\omega_c}{ne^2} = \frac{1}{nec}B. \tag{2.11}$$

The Hall resistivity is therefore linear in B and independent of both mass and relaxation time; more accurately, since we may consider even carriers of *positive* charge e ("holes"), its sign does depend on the carrier charge. Notice also that in the nondissipative regime ($\tau \gg 1/\omega_c$) both ρ_{xx} and σ_{xx} vanish.

If we write n as N/A (number of carriers per unit area), then

$$\rho_{xy} = \frac{1}{nec}B = \frac{\phi}{Nec}, \tag{2.12}$$

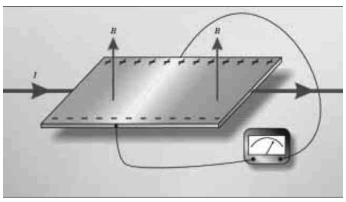

Figure 2.5: Hall effect in a 2d system. The $\mathbf{E}$ field is applied along x, while the $\mathbf{B}$ field is along z. The system is shorted in the y direction; the current $\mathbf{j}$ has both longitudinal (x) and transverse, a.k.a. Hall (y) components.

where $\phi = AB$ is the magnetic flux through area A. Although we are still at a purely classical level, it is instructive to multiply and divide ρ_{xy} by h. We may thus identically write

$$\rho_{xy} = \frac{1}{\nu}\frac{h}{e^2}, \quad \nu = \frac{N\phi_0}{\phi}. \tag{2.13}$$

Here $\phi_0 = hc/e$ is the flux quantum, as defined above. The dimensionless quantity ν, called the filling factor, equals the ratio between the number of electrons N and the number of flux quanta ϕ/ϕ_0. Eq. (2.13) expresses the transverse resistivity in terms of the natural resistance unit h/e^2. Since 1990 this is a new metrology standard, accurate to more than nine figures: 1 klitzing $= h/e^2 = 25812.807557(18)$ ohm.

In 2d resistance and resistivity have the same dimensions, and coincide in the transverse case. We write therefore the Hall resistance as

$$V_y/I_x = R_{\mathrm{H}} \equiv R_{xy} = \frac{1}{\nu}\frac{h}{e^2}. \tag{2.14}$$

Upon obvious dimensionality arguments, even in the quantum case the Hall resistance can be written in this way; but then the concentration- and B-dependence of ν are expected to be very different from the simple monothonical form of Eq. (2.13).

2.4.2 Landau levels

In quantum mechanics, the Schrödinger equation for an electron in 2d subject to a perpendicular **B** field (and in a flat potential) can be exactly dealt with, both in the Landau gauge and in the central gauge. The spectrum is discrete $\varepsilon_n = (n + \frac{1}{2})\hbar\omega_{\mathrm{c}}$. We define the magnetic length as $\ell = (\hbar c/eB)^{1/2}$; it diverges in the zero-field limit, and is of the order of 100 angstrom in a typical quantum Hall experiment. In the Landau gauge ($A_x = By, A_y = 0$) the eigenfunctions with energy n are

$$\psi_{nk}(x, y) \propto \mathrm{e}^{ikx}\chi_n(y - \ell^2 k), \tag{2.15}$$

where $\chi_n(y)$ are harmonic oscillator eigenfunctions with frequency ω_{c}. Each LL is infinitely degenerate (one eigenfunction for each k). For a system of area A, the number of states in each level is $\mathcal{N} = A/(2\pi\ell^2)$; this has a simple form in terms of the magnetic flux ϕ through area A: $\mathcal{N} = \phi/\phi_0$.

If we now consider N noninteracting electrons, the lowest LL is completely filled when $N = \mathcal{N}$; more generally, one expects a periodicity in the filling factor $\nu = N/\mathcal{N} = N\phi/\phi_0$, whenever ν crosses an integer value, in most physical properties.

2.4.3 The experiment

The Hall resistance of a noninteracting 2d electron gas has been computed quantum-mechanically by Ando in 1974 [38]. The result, when expressed as in Eq. (2.14) showed indeed oscillations in ν with integer period. The experiment, performed by von Klitzing and collaborators in 1980 [37], provided qualitatively different and very surprising results, shown in Fig 2.6. The discovery of the quantum Hall effect triggered a revolution with far reaching consequences in electronic structure theory at large; Klaus von Klitzing was awarded the Nobel prize in 1985.

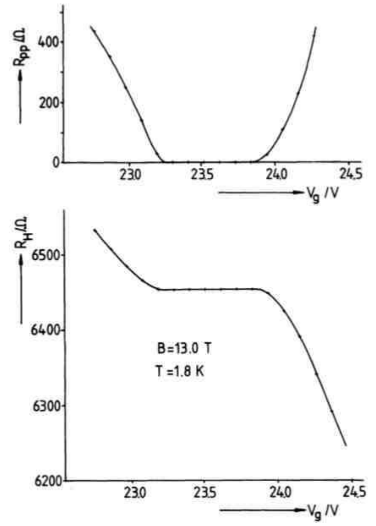

Figure 2.6: The original figure from von Klitzing et al. Ref. [37]. The gate voltage V_g was supposed to control the carrier density. Instead, the Hall resistance is quantized and insensitive to V_g over a large interval; over the same interval, the longitudinal resistance vanishes.

In the original experiment, the 2d electrons were confined by a MOSFET (metal-oxide-semi-conductor field-effect transistor); later, higher mobilities were obtained at semiconductor heterojunctions. Fig. 2.6 shows a very robust plateau, where $R_{xx} = 0$ and $R_{xy} = 6453.3 \pm 0.1$, corresponding to the filling factor $\nu = 4$. The accuracy in the quantized R_{xy} value is clearly far beyond the experimental control of the carrier concentration and of the $\mathbf{B}$ field uniformity over the sample. A novel, qualitatively different, state of matter was discovered. In modern jargon, the plateaus are "topologically protected".

A modern realization of the integer quantum Hall effect is shown in Fig. 2.7, where ρ_{xx} and ρ_{xy} are plotted as a function of the magnetic field. The plateau quantization is accurate to nine figures. The 2d electron gas is typically confined at a GaAs/GaAlAs heterojunction. The $\nu = 1$ value is achieved above $\simeq 10$ tesla; at low field (high ν) the system becomes dissipative ($\rho_{xx} > 0$), while

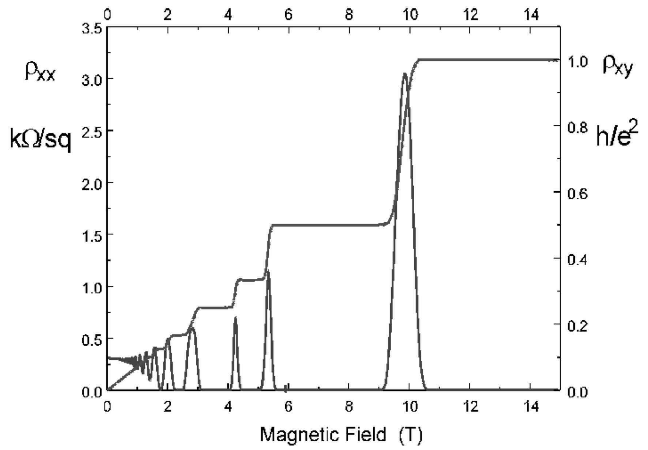

Figure 2.7: A modern realization of the integer quantum Hall effect

the classical linear behavior of ρ_{xy} is recovered; the slope depends on electron concentration n.

2.4.4 Early theoretical interpretation

The breakthrough paper, by Laughlin, appeared as early as 1981 [39]. This is a remarkably concise paper (two pages) which, in retrospect, is based on topological arguments. One key ingredient of the theory is *disorder*: in fact, the quantum Hall effect becomes less spectacular for very "clean" samples, while some "dirtyness" enhances the effect.

Laughlin devised a Gedankenexperiment based on the setup shown in Fig. 2.8. The corresponding 2d Schrödinger Hamiltonian in the Landau gauge is

$$H(\varphi) = \frac{1}{2m}\left[\left(p_x + \frac{e}{c}A_x\right)^2 + p_y^2\right] + eEy + \mathscr{V}(x,y), \tag{2.16}$$

where E is the field across the ribbon, and $\mathscr{V}$ is an arbitrary substrate potential. The addition of a constant vector potential along x, $A_x \to A_x + \Delta A$, in Eq. (2.16) corresponds to threading a flux $\varphi = L\,\Delta A$ through the cylinder; we use the symbol φ, not to be confused with the real magnetic flux ϕ normal to the surface.

Similarly to what discussed in Sec. 1.6, the eigenvalues acquire a φ dependence. According to Eq. (1.21), if $\varepsilon_n(\varphi)$ is the n-th eigenvalue the current transported by the corresponding eigenstate is $I_x = -c\,\partial\varepsilon_n(\varphi)/\partial\varphi$. For an independent-particle system with N carriers the current is thus

$$I_x = -c\frac{\partial U(\varphi)}{\partial\varphi}, \tag{2.17}$$

where $U(\varphi)$ is the total energy of the system. Implicitly, we are assuming a dissipationless system.

The expression in Eq. (2.17) for the current is remarkably simple, general, and robust: it does not depend on the substrate potential $\mathscr{V}(x,y)$, nor the number N of carriers, and not even on their mass m. But for a disordered potential the eigenstates come in two kinds: localized and extended. The latter ones are phase-coherent round the loop, while the former are exponentially localized for $L \to \infty$. The localized states are insensitive to the flux insertion (like the OBCs eigenstates in Sec. 1.6), and the whole current is carried by the delocalized ones. Therefore Eq. (2.17) provides a nonzero result insofar at least one of the occupied eigenstates in the disordered sample is extended,

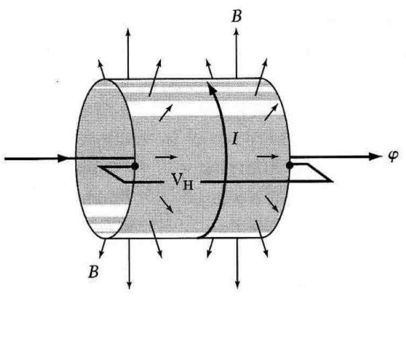

Figure 2.8: Geometry for Laughlin's Gedankenexperiment. A 2d channel is bent into a loop of circumference L, and a magnetic $\mathbf{B}$ field of constant magnitude pierces the cylinder normal to the surface. A current $I \equiv I_x$ circles the loop; $V_{\mathrm{H}} \equiv V_y$ is the Hall voltage. The loop may be threaded by the inaccessible flux φ.

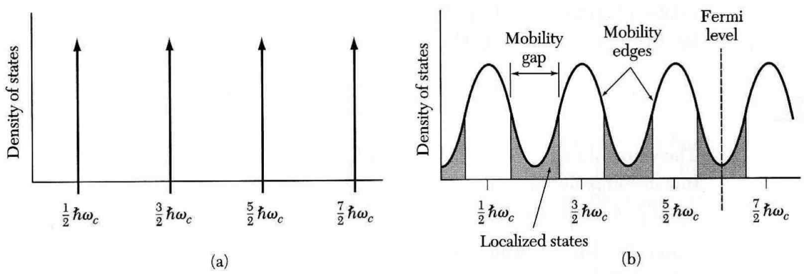

Figure 2.9: The density of states for a 2d system of noninteracting carriers. (a) Clean system, with zero substrate potential. (b) Actual sample, in presence of substrate disorder and impurities.

i.e. phase-coherent round the loop; besides this, the number and nature of the current-carrying stated is irrelevant. It is therefore crucial to address the nature of the single-particle eigenstates in a quantum Hall sample. For a clean sample (flat substrate) the LLs are sharp, all eigenstates are extended (in the Landau gauge), and the density of states is a series of delta functions, shown in Fig. 2.9 (a); the weight of each delta is ϕ/ϕ_0. In presence of disorder, the deltas broaden into alternating bands of localized and extended states, as sketched in Fig. 2.9 (b).

The electron fluid is in the quantum Hall regime whenever the Fermi level falls in a region of localized states. Therefore $\sigma_{xx} = 0$ (the fluid is a "quantum Hall insulator"), and $\rho_{xx} = 0$ (transport is dissipationless).

We now imagine to adiabatically increase the vector potential by an amount $\Delta A = \phi_0/L$, where ϕ_0 is a flux quantum: all of the current-carrying states are mapped back into themselves, while the localized ones are unaffected. Hence the ground state has the same energy; nonetheless Eq. (2.17) implies $U(\varphi + \phi_0) - U(\varphi) \simeq -\phi_0 I_x/c \neq 0$. This is only possible if an integer number of electrons is transferred from one cylinder edge to the other, each of them contributing the energy eV_x. If we call $-\nu$ such integer number, the relationship is then

$$\phi_0 I_x/c = \nu e V_y; \qquad R_{\mathrm{H}} = V_y/I_x = \frac{\phi_0}{\nu c e} = \frac{1}{\nu}\frac{h}{e^2}. \tag{2.18}$$

The flux φ acts therefore as a charge pump; the pump cycle is one flux quantum ϕ_0.

Ideally, the sample ground state can be continuously "deformed" from dirty to clean. Insofar as the Fermi level stays is in a region of nonconducting states, the (topological) integer ν cannot change, even if the number of current carrying states does obviously change. The identification of ν with the number of filled LLs comes from the clean-sample limit, which is exactly soluble. Setting $\mathscr{V}(x, y) = 0$ the eigenfunctions of Eq. (2.16) are

$$\psi_{nk}(x, y) = \mathrm{e}^{ikx}\chi_n(y - y_0), \quad y_0 = \ell^2 k - \frac{cE}{\omega_{\mathrm{c}} B}. \tag{2.19}$$

For a finite L, the allowed k's are integer multiples of $2\pi/L$ and the corresponding centers y_0 are spaced by $2\pi\ell^2/L = L\phi_0/\phi$. Threading a flux φ shifts y_0 linearly in φ; when φ equals one flux quantum each eigenfunction goes over to

the next. Therefore one carrier is shifted for each n; the integer index ν measures therefore the number of occupied LLs. Similar arguments can be reformulated in different gauges and in different geometries [40, 41].

Chapter 3

Berry-ology

Berry-ology is used as synonymous for geometry in nonrelativistic quantum mechanics. Topological (i.e. quantized) quantities are defined via geometrical quantities, analogously to what we did above in Sec. 1.2. But many important quantities (notably the Berry phase) are merely geometrical.

Let us assume that a generic time-independent quantum Hamiltonian has a parametric dependence. The Schrödinger equation is

$$H(\boldsymbol{\xi})|\Psi(\boldsymbol{\xi})\rangle = E(\boldsymbol{\xi})|\Psi(\boldsymbol{\xi})\rangle, \tag{3.1}$$

where the d-dimensional real parameter $\boldsymbol{\xi}$ is defined in a suitable domain of $\mathbb{R}^d$: a 2d $\boldsymbol{\xi}$ has been chosen for display in Fig. 3.1. In most of this Section we discuss the most general case, and therefore we do not specify which quantum system is described by this Hamiltonian, nor what the physical meaning of the parameter $\boldsymbol{\xi}$ is.

In the subsequent Chapters $|\Psi(\boldsymbol{\xi})\rangle$ will be identified with either a single-particle wavefunction (a.k.a. orbital) or a many-electron wavefunction. As for the parameter $\boldsymbol{\xi}$, it may be a nuclear coordinate, a phase angle, a magnetic flux, a Bloch vector, a momentum, and more: it could therefore have various dimensions. Sometimes, the parameter $\boldsymbol{\xi}$ is called the "slow variable", while the electronic coordinates are the "fast variable". At the end of this Chapter, Sec. 3.9, we address the special case where the parameter $\boldsymbol{\xi}$ is identified with a Bloch vector.

The state vectors $|\Psi(\boldsymbol{\xi})\rangle$ are all supposed to be normalized and to reside in the same Hilbert space: this amounts to saying that the wavefunctions are supposed to obey $\boldsymbol{\xi}$-independent boundary conditions. We focus on the ground state $|\Psi_0(\boldsymbol{\xi})\rangle$, and we assume it to be nondegenerate for $\boldsymbol{\xi}$ in some domain of $\mathbb{R}^d$.

Any quantum mechanical state vector is arbitrary by a constant phase factor. Here we refer to choosing this phase as to the choice of the gauge. The semantic is a bit ambiguous. In presence of magnetic fields, we may change the magnetic gauge: this changes the Hamiltonian and the eigenfunctions. Once the magnetic gauge—hence the Hamiltonian—is fixed, we still remain with the phase arbitrariness referred to above. All measurable quantities (e.g. expectation values) are obviously gauge-invariant (in both senses), but the reverse is also true: all gauge-invariant properties are—at least in principle—measurable.

It is expedient to define the ground-state projector (a.k.a. density matrix)

and its complement, i.e.

$$\hat{P}(\boldsymbol{\xi}) = |\Psi_0(\boldsymbol{\xi})\rangle\langle\Psi_0(\boldsymbol{\xi})|; \quad \hat{Q}(\boldsymbol{\xi}) = \hat{1} - \hat{P}(\boldsymbol{\xi}). \tag{3.2}$$

Both $\hat{P}(\boldsymbol{\xi})$ and $\hat{Q}(\boldsymbol{\xi})$ are gauge-invariant (for a fixed Hamiltonian).

3.1 Phases and distances

We define the *phase difference* between the ground eigenstates at two different $\boldsymbol{\xi}$ points in the most natural way:

$$\mathrm{e}^{-i\Delta\varphi_{12}} = \frac{\langle\Psi_0(\boldsymbol{\xi}_1)|\Psi_0(\boldsymbol{\xi}_2)\rangle}{|\langle\Psi_0(\boldsymbol{\xi}_1)|\Psi_0(\boldsymbol{\xi}_2)\rangle|} \; ; \tag{3.3}$$

$$\Delta\varphi_{12} = - \text{ Im log } \langle\Psi_0(\boldsymbol{\xi}_1)|\Psi_0(\boldsymbol{\xi}_2)\rangle \; . \tag{3.4}$$

For any given choice of the two states, Eqs. (3.3) and (3.4) provide a $\Delta\varphi_{12}$ which is unique modulo 2π, except in the very special case that the states are orthogonal. However, it is also clear that such $\Delta\varphi_{12}$ is gauge-dependent and *cannot* have, by itself, any physical meaning.

The distance between quantum states has been defined by Bures [42] as:

$$D^2_{12} = 1 - |\langle\Psi_0(\boldsymbol{\xi}_1)|\Psi_0(\boldsymbol{\xi}_2)\rangle|^2. \tag{3.5}$$

Such distance fulfills the familiar axioms from calculus textbooks; it vanishes when the two states physically coincide (i.e. independently of the phase factor), and is maximum (equal to one) when the states are orthogonal. At variance with $\Delta\varphi_{12}$ defined above, the Bures distance is gauge-invariant, and can be explicitly expressed in terms of ground-state projectors

$$D^2_{12} = 1 - \text{Tr } \{\hat{P}(\boldsymbol{\xi}_1)\hat{P}(\boldsymbol{\xi}_2)\}, \tag{3.6}$$

where "Tr" is the trace over the Hilbert space.

3.2 Berry phase

We have already observed that the phase difference $\Delta\varphi_{12}$ between any two states is gauge-dependent and cannot have any physical meaning. Matters are quite different when we consider the *total* phase difference along a closed path which joins several points in a given order, as shown in Fig. 3.2:

$$\begin{aligned} \gamma &= \Delta\varphi_{12} + \Delta\varphi_{23} + \Delta\varphi_{34} + \Delta\varphi_{41} \\ &= - \text{ Im log } \langle\Psi_0(\boldsymbol{\xi}_1)|\Psi_0(\boldsymbol{\xi}_2)\rangle\langle\Psi_0(\boldsymbol{\xi}_2)|\Psi_0(\boldsymbol{\xi}_3)\rangle \times \\ &\times \; \langle\Psi_0(\boldsymbol{\xi}_3)|\Psi_0(\boldsymbol{\xi}_4)\rangle\langle\Psi_0(\boldsymbol{\xi}_4)|\Psi_0(\boldsymbol{\xi}_1)\rangle. \end{aligned} \tag{3.7}$$

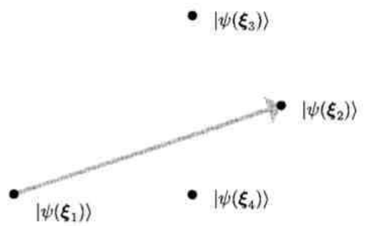

Figure 3.1: State vectors in the two-dimensional $\boldsymbol{\xi}$-space. The phase difference between two of them is defined as $\mathrm{e}^{-i\Delta\varphi_{12}} = \frac{\langle\Psi_0(\boldsymbol{\xi}_1)|\Psi_0(\boldsymbol{\xi}_2)\rangle}{|\langle\Psi_0(\boldsymbol{\xi}_1)|\Psi_0(\boldsymbol{\xi}_2)\rangle|}$, and their distance as $D^2_{12} = 1 - |\langle\Psi_0(\boldsymbol{\xi}_1)|\Psi_0(\boldsymbol{\xi}_2)\rangle|^2$.

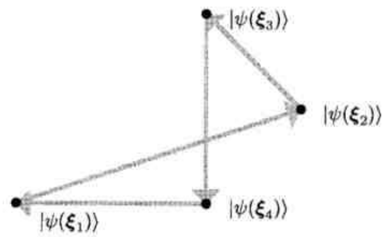

Figure 3.2: A closed path joining four states in $\boldsymbol{\xi}$-space.

It is now clear that all the gauge-arbitrary phases cancel in pairs, such as to make the overall phase γ a gauge–invariant quantity. The above simple–minded algebra leads to a result of overwhelming physical importance: in fact, a gauge–invariant quantity is potentially a physical observable. In essence, this is the revolutionary message of Berry's celebrated paper, appeared in 1984 [27, 43].

Next we consider a smooth closed curve C in the parameter domain, such as in Fig. 3.3, and we discretize it with a set of points on it. Using Eq. (3.3), we write the phase difference between any two contiguous points as

$$\mathrm{e}^{-i\Delta\varphi} = \frac{\langle\Psi_0(\boldsymbol{\xi})|\Psi_0(\boldsymbol{\xi}+\Delta\boldsymbol{\xi})\rangle}{|\langle\Psi_0(\boldsymbol{\xi})|\Psi_0(\boldsymbol{\xi}+\Delta\boldsymbol{\xi})\rangle|}. \tag{3.8}$$

If we further assume that the gauge is so chosen that the phase varies in a *differentiable* way along the path, then from Eq. (3.8) we get to leading order in $\Delta\boldsymbol{\xi}$:

$$-i\Delta\varphi \simeq \langle\Psi_0(\boldsymbol{\xi})|\nabla_{\boldsymbol{\xi}}\Psi_0(\boldsymbol{\xi})\rangle \cdot \Delta\boldsymbol{\xi}. \tag{3.9}$$

In the limiting case of a set of points which becomes dense on the continuous path, the total phase difference γ converges to a circuit integral:

$$\gamma = \sum_{s=1}^{M} \Delta\varphi_{s,s+1} \;\longrightarrow\; \oint_C \boldsymbol{\mathcal{A}}(\boldsymbol{\xi}) \cdot d\boldsymbol{\xi}, \tag{3.10}$$

where $\boldsymbol{\mathcal{A}}(\boldsymbol{\xi})$ is called the Berry connection:

$$\boldsymbol{\mathcal{A}}(\boldsymbol{\xi}) = i\,\langle\Psi_0(\boldsymbol{\xi})|\nabla_{\boldsymbol{\xi}}\Psi_0(\boldsymbol{\xi})\rangle. \tag{3.11}$$

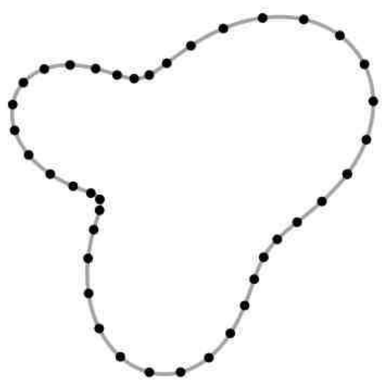

Figure 3.3: A smooth closed curve C in $\boldsymbol{\xi}$-space, and its discretization.

Since the state vectors are assumed to be normalized at any $\boldsymbol{\xi}$, the connection is *real*; we can therefore equivalently write

$$\boldsymbol{\mathcal{A}}(\boldsymbol{\xi}) = -\text{Im}\,\langle \Psi_0(\boldsymbol{\xi}) | \nabla_{\boldsymbol{\xi}} \Psi_0(\boldsymbol{\xi}) \rangle. \tag{3.12}$$

A number of manifestations of the Berry phase occurring in molecular and condensed matter phenomena will be discussed in the present Notes. Many reviews papers; here we quote Refs. [43, 44, 45, 3, 12].

At this point it is also worth emphasizing that in computational physics there are no derivatives. The ground state $\Psi_0(\boldsymbol{\xi})$ is generally found by diagonalizing a matrix on a finite set of $\boldsymbol{\xi}$ points, and the phase (i.e. the gauge) is chosen by the diagonalization routine; the phase is therefore nonsmooth and possibly even random. The discrete form in Eq. (3.11) at finite M is the one universally used in numerical work; it is clearly unaffected by any erratic phase factor.

3.3 Connection and curvature

The loop integral of the Berry connection (i.e. the Berry phase γ) is non trivial in two cases: either the curl of $\boldsymbol{\mathcal{A}}(\boldsymbol{\xi})$ is nonzero, or the curl is zero but the curve C is not in a simply connected domain. In the former case, we can invoke Stokes theorem; the formulation is very simple when $\boldsymbol{\xi}$ is a 3d parameter. If C is the boundary of a surface Σ (i.e. $C \equiv \partial\Sigma$), and the curl of $\boldsymbol{\mathcal{A}}(\boldsymbol{\xi})$ is regular on Σ, then Stokes' theorem reads

$$\gamma = \oint_{\partial\Sigma} \boldsymbol{\mathcal{A}}(\boldsymbol{\xi}) \cdot d\boldsymbol{\xi} = \int_{\Sigma} \boldsymbol{\Omega}(\boldsymbol{\xi}) \cdot \mathbf{n}\; d\sigma, \tag{3.13}$$

where $\boldsymbol{\Omega}$ is the Berry curvature, defined as

$$\begin{aligned} \boldsymbol{\Omega}(\boldsymbol{\xi}) &= \nabla_{\boldsymbol{\xi}} \times \boldsymbol{\mathcal{A}}(\boldsymbol{\xi}) = -\text{Im}\,\langle \nabla_{\boldsymbol{\xi}} \Psi_0(\boldsymbol{\xi})| \times |\nabla_{\boldsymbol{\xi}} \Psi_0(\boldsymbol{\xi})\rangle \\ &= i \langle \nabla_{\boldsymbol{\xi}} \Psi_0(\boldsymbol{\xi})| \times |\nabla_{\boldsymbol{\xi}} \Psi_0(\boldsymbol{\xi})\rangle, \end{aligned} \tag{3.14}$$

with the usual meaning of the cross product between three-component bra and ket states. Equation (3.13) may be spelled out by saying that the curvature is the Berry phase per unit area of Σ.

For $d \neq 3$ the Berry curvature is conveniently written as the $d \times d$ antisymmetric matrix

$$\boldsymbol{\Omega}_{\alpha\beta}(\boldsymbol{\xi}) = -2\,\text{Im}\,\langle \partial_\alpha \Psi_0(\boldsymbol{\xi}) | \partial_\beta \Psi_0(\boldsymbol{\xi}) \rangle; \tag{3.15}$$

Greek subscripts are Cartesian coordinates throughout, and $\partial_\alpha = \partial/\partial\xi_\alpha$. The Stokes theorem can still be applied, generalizing Eq. (3.13) to

$$\gamma = \frac{1}{2} \int_{\Sigma} d\xi^\alpha \wedge d\xi^\beta\; \boldsymbol{\Omega}_{\alpha\beta}(\boldsymbol{\xi}). \tag{3.16}$$

The Berry connection is also known as "gauge potential", and the Berry curvature as "gauge field" [45]. It is worth pointing out that the former is gauge-dependent, while the latter is gauge-invariant and therefore corresponds in general to a measurable quantity, even before any integration. The two quantities play (in $\boldsymbol{\xi}$-space) a similar role as the vector potential and the magnetic field in elementary magnetostatics: $\mathbf{A}(\mathbf{r})$ is gauge-dependent, nonmeasurable; $\mathbf{B}(\mathbf{r}) = \nabla_{\mathbf{r}} \times \mathbf{A}(\mathbf{r})$ is gauge-invariant, measurable.

The Berry phase γ, defined as the integral over a closed curve C of the connection, is gauge invariant only modulo 2π. This indeterminacy is resolved by Eqs. (3.13) and (3.16) whenever the curve C is the boundary $\partial\Sigma$ of a surface Σ where the curvature is regular. In fact, the curvature is gauge-invariant and has no modulo 2π indeterminacy.

3.4 Chern number

The rhs of Eqs. (3.13) and (3.16) is the flux of the Berry curvature on the surface Σ; such flux remains meaningful even on a closed surface (e.g. a sphere or a torus), in which case $\partial\Sigma$ is the empty set. The key result is that such an integral is quantized. Here we limit ourselves to 3d, where we identify Σ with the sphere S^2 (Fig. 3.4):

$$\frac{1}{2\pi}\int_{S^2} \mathbf{\Omega}(\boldsymbol{\xi})\cdot\mathbf{n}\, d\sigma = C_1; \tag{3.17}$$

C_1 is an integer $\in \mathbb{Z}$, called Chern number of the first class.

The proof is based on a similar algebra as for Dirac's theory of the magnetic monopole [44, 46]. The theorem goes sometimes under the name of Gauss-Bonnet-Chern theorem; the analogy with Eq. (1.4) is perspicuous. A specific example is dealt with in detail in Sec. 4.1.

The curvature is regular (and divergence-free) on the closed surface S^2; the lhs of Eq. (3.17) is the flux of $\mathbf{\Omega}(\boldsymbol{\xi})$ across S^2. The integrand $\mathbf{\Omega}(\boldsymbol{\xi})$ is the curl of the connection $\boldsymbol{\mathcal{A}}(\boldsymbol{\xi})$; the latter in general *cannot* be defined as a single-valued function globally on S^2, but only on *patches* of it [44, 46]. To fix the ideas, suppose that $\mathbf{\Omega}(\boldsymbol{\xi})$ is singular at $\boldsymbol{\xi} = 0$, and that S^2 is the spherical surface centered at the origin (Fig. 3.4). We cut this surface at the equator $\xi_z = 0$ and we consider the flux across the two open surfaces:

$$\int_{S^2} \mathbf{\Omega}(\boldsymbol{\xi})\cdot\mathbf{n}\, d\sigma = \int_{S_+} \mathbf{\Omega}(\boldsymbol{\xi})\cdot\mathbf{n}\, d\sigma + \int_{S_-} \mathbf{\Omega}(\boldsymbol{\xi})\cdot\mathbf{n}\, d\sigma. \tag{3.18}$$

We notice that $\partial S_+ = \partial S_- = C$, but the surface normals $\mathbf{n}$ have opposite orientations. From Stokes theorem, Eq. (3.13), we get:

$$\int_{S_\pm} \mathbf{\Omega}(\boldsymbol{\xi})\cdot\mathbf{n}\, d\sigma = \pm \oint_C \boldsymbol{\mathcal{A}}_\pm(\boldsymbol{\xi})\cdot d\boldsymbol{\xi} \tag{3.19}$$

$$\int_{S^2} \mathbf{\Omega}(\boldsymbol{\xi})\cdot\mathbf{n}\, d\sigma = \oint_C \boldsymbol{\mathcal{A}}_+(\boldsymbol{\xi})\cdot d\boldsymbol{\xi} - \oint_C \boldsymbol{\mathcal{A}}_-(\boldsymbol{\xi})\cdot d\boldsymbol{\xi}\,. \tag{3.20}$$

The two upper and lower Berry connections $\boldsymbol{\mathcal{A}}_\pm(\boldsymbol{\xi})$ may only differ by a gauge transformation; the rhs of Eq. (3.20) is the difference of two Berry phases on

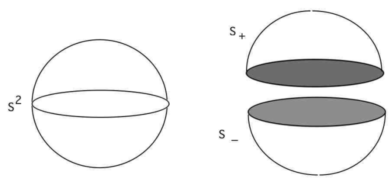

Figure 3.4: A sphere cut at the equator in two hemispheres.

the same path and is necessarily a multiple of 2π. This concludes the proof of Eq. (3.17).

We emphasize that the Chern number is a robust topological invariant of the wavefunction, and is at the origin of observable effects. The Chern number made its first appearance in electronic structure in 1982, in the famous TKNN paper about the quantum Hall effect [47] (Sec. 4.2.3). Nowadays more complex topological invariants are in fashion, and they characterize a completely novel class of insulators, called "topological insulators" [10, 14, 48, 49, 50, 51, 52, 53].

3.5 Metric

Starting from Eq. (3.5), the infinitesimal distance is

$$D^2_{\boldsymbol{\xi},\boldsymbol{\xi}+d\boldsymbol{\xi}} = \sum_{\alpha,\beta=1}^{d} g_{\alpha\beta}(\boldsymbol{\xi}) d\xi_\alpha d\xi_\beta, \tag{3.21}$$

where the metric tensor is easily shown to be

$$\begin{aligned} g_{\alpha\beta}(\boldsymbol{\xi}) &= \text{Re } \langle \partial_\alpha \Psi_0(\boldsymbol{\xi}) | \partial_\beta \Psi_0(\boldsymbol{\xi}) \rangle \\ &- \langle \partial_\alpha \Psi_0(\boldsymbol{\xi}) | \Psi_0(\boldsymbol{\xi}) \rangle \langle \Psi_0(\boldsymbol{\xi}) | \partial_\beta \Psi_0(\boldsymbol{\xi}) \rangle \\ &= \text{Re } \langle \partial_\alpha \Psi_0(\boldsymbol{\xi}) | \hat{Q}(\boldsymbol{\xi}) | \partial_\beta \Psi_0(\boldsymbol{\xi}) \rangle; \end{aligned} \tag{3.22}$$

the projector $\hat{Q}(\boldsymbol{\xi})$ is the same as defined in Eq. (3.2). This quantum metric tensor was first proposed by Provost and Vallee in 1980 [54].

At this point we may compare Eq. (3.22) to Eq. (3.15), noticing that the insertion of $\hat{Q}(\boldsymbol{\xi})$ is irrelevant in the latter, i.e.

$$\boldsymbol{\Omega}_{\alpha\beta}(\boldsymbol{\xi}) = -2 \,\text{Im } \langle \partial_\alpha \Psi_0(\boldsymbol{\xi}) | \hat{Q}(\boldsymbol{\xi}) | \partial_\beta \Psi_0(\boldsymbol{\xi}) \rangle. \tag{3.23}$$

It is therefore clear that $g_{\alpha\beta}$ and $\boldsymbol{\Omega}_{\alpha\beta}$ are, apart for a trivial -2 factor, the real (symmetric) and the imaginary (antisymmetric) parts of the same tensor, which we are going to call $\mathcal{F}_{\alpha\beta}$ in the following:

$$\mathcal{F}_{\alpha\beta}(\boldsymbol{\xi}) = \langle \partial_\alpha \Psi_0(\boldsymbol{\xi}) | \hat{Q}(\boldsymbol{\xi}) | \partial_\beta \Psi_0(\boldsymbol{\xi}) \rangle. \tag{3.24}$$

The metric-curvature tensor $\mathcal{F}_{\alpha\beta}$ is gauge-invariant. A compact equivalent expression is

$$\mathcal{F}_{\alpha\beta}(\boldsymbol{\xi}) = \text{Tr } \{ \partial_\alpha \hat{P}(\boldsymbol{\xi}) \hat{Q}(\boldsymbol{\xi}) \partial_\beta \hat{P}(\boldsymbol{\xi}) \}, \tag{3.25}$$

manifestly gauge-invariant and Hermitian.

3.6 Sum over states

The $\boldsymbol{\xi}$-derivatives entering many of the previous equations, e.g. Eq. (3.24), can be expressed starting from perturbation theory:

$$\begin{aligned} & |\Psi_0(\boldsymbol{\xi} + \Delta\boldsymbol{\xi})\rangle - |\Psi_0(\boldsymbol{\xi})\rangle \\ \simeq \; & \sum_{n\neq 0}{}' |\Psi_n(\boldsymbol{\xi})\rangle \frac{\langle \Psi_n(\boldsymbol{\xi}) | \, [\, H(\boldsymbol{\xi} + \Delta\boldsymbol{\xi}) - H(\boldsymbol{\xi}) \,] \, | \Psi_0(\boldsymbol{\xi}) \rangle}{E_0(\boldsymbol{\xi}) - E_n(\boldsymbol{\xi})}; \end{aligned} \tag{3.26}$$

$$|\partial_\alpha \Psi_0(\boldsymbol{\xi})\rangle = {\sum_{n\neq 0}}' |\Psi_n(\boldsymbol{\xi})\rangle \frac{\langle \Psi_n(\boldsymbol{\xi})|\partial_\alpha H(\boldsymbol{\xi})|\Psi_0(\boldsymbol{\xi})\rangle}{E_0(\boldsymbol{\xi}) - E_n(\boldsymbol{\xi})}. \tag{3.27}$$

These seemingly obvious and innocent formulae need some caveat. It is clear that inserting Eq. (3.27) into the Berry connection, Eqs. (3.11) and (3.12), we get a vanishing result for any $\boldsymbol{\xi}$. This happens because the simple expression of Eq. (3.26) corresponds to a very specific gauge choice (called the parallel-transport gauge [3]); multiplying the rhs by a $\boldsymbol{\xi}$-dependent phase factor is legitimate, and must not modify any physical result, while the Berry connection is instead affected. Nonetheless, since our $\mathcal{F}_{\alpha\beta}(\boldsymbol{\xi})$ is a gauge-invariant quantity, we may safely evaluate it in any gauge, including the parallel-transport gauge, implicit in Eq. (3.27). The result is

$$\begin{aligned} &\mathcal{F}_{\alpha\beta}(\boldsymbol{\xi}) \\ &= {\sum_{n\neq 0}}' \quad \frac{\langle \Psi_0(\boldsymbol{\xi})|\partial_\alpha H(\boldsymbol{\xi})|\Psi_n(\boldsymbol{\xi})\rangle\langle \Psi_n(\boldsymbol{\xi})|\partial_\beta H(\boldsymbol{\xi})|\Psi_0(\boldsymbol{\xi})\rangle}{[E_0(\boldsymbol{\xi}) - E_n(\boldsymbol{\xi})]^2}. \end{aligned} \tag{3.28}$$

This expression shows explicitly that both the curvature and the metric are ill defined and singular wherever the ground state is degenerate with the first excited state. Indeed, this is the main reason why the domain may happen not to be simply connected.

3.7 Time-reversal and inversion symmetries

According to Eq. (3.25) the ground-state projector uniquely determines the curvature

$$\boldsymbol{\Omega}_{\alpha\beta}(\boldsymbol{\xi}) = -2\,\mathrm{Im}\ \mathrm{Tr}\ \{\partial_\alpha \hat{P}(\boldsymbol{\xi})\hat{Q}(\boldsymbol{\xi})\partial_\beta \hat{P}(\boldsymbol{\xi})\}. \tag{3.29}$$

It is therefore expedient to analyze the symmetries of the ground-state projector $\hat{P}(\boldsymbol{\xi})$, which coincide with the symmetries of the Hamiltonian; these in turn depend on the nature of the parameter $\boldsymbol{\xi}$. We only address spinless electrons (no spin-orbit coupling).

When $\boldsymbol{\xi}$ is even under time reversal (like e.g. a nuclear coordinate), then time-reversal invariance implies that both $H(\boldsymbol{\xi})$ and $\hat{P}(\boldsymbol{\xi})$ are real for any $\boldsymbol{\xi}$, and Eq. (3.29) warrants that the curvature is everywhere vanishing. The Berry phase γ can be nonzero (modulo 2π) only if the curve C loops around a singularity or, more generally, it does not lie in a simply connected domain; the only allowed values are $\gamma = 0$ or $\gamma = \pi$.

When instead $\boldsymbol{\xi}$ is odd under time-reversal (like e.g. a momentum) then time-reversal symmetry requires $\hat{P}(-\boldsymbol{\xi}) = \hat{P}^*(\boldsymbol{\xi})$, therefore

$$\boldsymbol{\Omega}_{\alpha\beta}(-\boldsymbol{\xi}) = -\boldsymbol{\Omega}_{\alpha\beta}(\boldsymbol{\xi}). \tag{3.30}$$

The Berry phase along an inversion-symmetric path vanishes; the Chern number vanishes as well.

Next we switch to inversion symmetry. When evaluating any $\boldsymbol{\xi}$-dependent matrix-element (or trace), inversion of the coordinates at fixed $\boldsymbol{\xi}$ is equivalent to keeping the coordinates fixed and inverting $\boldsymbol{\xi}$. This statement holds whether $\boldsymbol{\xi}$ is a coordinate or a momentum; both are in fact odd under inversion. Therefore inversion symmetry implies $\hat{P}(-\boldsymbol{\xi}) = \hat{P}(\boldsymbol{\xi})$, and

$$\boldsymbol{\Omega}_{\alpha\beta}(-\boldsymbol{\xi}) = \boldsymbol{\Omega}_{\alpha\beta}(\boldsymbol{\xi}). \tag{3.31}$$

If both time-reversal and inversion symmetry are present, then the Berry curvature is everywhere vanishing. The Berry phase can be only zero or π; the latter case requires a domain which is not simply connected, as above.

Crucial to the above arguments is the fact that the double derivative appearing in Eq. (3.29) are even under either time-reversal or inversion. Summarizing the symmetry results, for the case where $\boldsymbol{\xi}$ is a momentum: a non vanishing Chern number can only occur in absence of time-reversal symmetry, but may occur even in inversion-symmetric cases.

3.8 NonAbelian case

Very soon after the appearance of Berry's milestone paper, Wilczek and Zee [55] introduced a many-state generalization. Suppose we do not focus on a single state in the Hilbert space (say the ground state), while we are interested instead in the behavior of n different states altogether, as a function of $\boldsymbol{\xi}$. The original motivation was the possibility of degeneracy amongst the n lowest eigenstates $|\psi_j(\boldsymbol{\xi})\rangle$ of a given Hamiltonian at some points of the path, and they formulated the theory under the hypothesis that these lowest states are never degenerate with the $n+1$-th. Strictly speaking, an Hamiltonian is unnecessary: one only needs to unambiguously identify an n-dimensional (n fixed) manifold of states, as function of the parameter $\boldsymbol{\xi}$. A gauge transformation is then a $n \times n$ unitary matrix; in the Abelian case, this matrix is a diagonal one, with unitary elements. The gauge invariant quantity which defines the manifold is the n-dimensional projector

$$P(\boldsymbol{\xi}) = \sum_{j=1}^{n} |\psi_j(\boldsymbol{\xi})\rangle\langle\psi_j(\boldsymbol{\xi})|. \tag{3.32}$$

Within loss of generality, we may address the case which is most interesting in electronic structure. We assume that the n states $|\psi_j(\boldsymbol{\xi})\rangle$ are spin orbitals, and that $|\Psi(\boldsymbol{\xi})\rangle$ is the many-body wavefunction built as their Slater determinant:

$$\Psi(\boldsymbol{\xi})\rangle = \frac{1}{\sqrt{N!}} |\psi_1(\boldsymbol{\xi})\psi_2(\boldsymbol{\xi}) \dots \psi_N(\boldsymbol{\xi})|. \tag{3.33}$$

We can therefore apply some of the results of the previous Sections. The many-body Berry phase is given in Eqs. (3.10) and (3.11), which we prefer to rewrite here in the form

$$\mathrm{e}^{-i\gamma} = \exp \oint_C \boldsymbol{\mathcal{A}}(\boldsymbol{\xi}) \cdot d\boldsymbol{\xi}, \tag{3.34}$$

where $\boldsymbol{\mathcal{A}}$ is the connection of the many-body wavefunction.

In the nonAbelian case the connection generalizes to a vector of $n \times n$ Hermitian matrices

$$\mathscr{A}_{jj'}(\boldsymbol{\xi}) = -\mathrm{Im}\ \langle\psi_j(\boldsymbol{\xi})|\nabla_{\boldsymbol{\xi}}\psi_{j'}(\boldsymbol{\xi})\rangle, \tag{3.35}$$

and the Berry phase factor of Eq. (3.34) generalizes to the unitary matrix

$$\mathrm{e}^{-i\Gamma} = \mathscr{P} \exp \oint_C \mathscr{A}(\boldsymbol{\xi}) \cdot d\boldsymbol{\xi}. \tag{3.36}$$

Here $\mathscr{P}$ is a path-ordering operator, owing to the noncommuting nature of the $\mathscr{A}$ matrices at different $\boldsymbol{\xi}$ in the nonAbelian case. The $\mathscr{P}$ operator has a

precise meaning when the series expansion of the exponential is considered. The discretization of Eq. (3.36) is rather straightforward and will not be discussed here [19].

While the single-state phase factor $e^{-i\gamma}$ is gauge-invariant, the Wilczek-Zee unitary matrix $e^{-i\Gamma}$ is only gauge-*covariant*, i.e. a gauge transformation yields a matrix *unitarily equivalent* to $e^{-i\Gamma}$; the gauge can be fixed by choosing the vectors $|\psi_j(\boldsymbol{\xi})\rangle$ at one point of the path. The eigenvalues γ_j of the Hermitian matrix Γ, defined modulo 2π, are gauge-invariant and in principle individually observable. It is easily proved that their sum, i.e. the trace of Γ, coincides with the Berry phase γ of the many-body wavefunction, Eq. (3.34) [19].

In order to write the metric-curvature tensor in the nonAbelian case, we start writing the many-body Abelian metric-curvature tensor of the Slater-determinant wavefunction, Eq. (3.24), as

$$\mathcal{F}_{\alpha\beta}(\boldsymbol{\xi}) = \langle \partial_\alpha \Psi_0(\boldsymbol{\xi}) | \partial_\beta \Psi_0(\boldsymbol{\xi}) \rangle - \langle \partial_\alpha \Psi_0(\boldsymbol{\xi}) | \hat{P}(\boldsymbol{\xi}) | \partial_\beta \Psi_0(\boldsymbol{\xi}) \rangle. \tag{3.37}$$

Next, we need now to explicitate the Cartesian indices α, β at the same time as the matrix indices j, j'. The nonAbelian metric-curvature tensor is the generalized form of Eq. (3.25), i.e.

$$\mathscr{F}_{\alpha\beta,jj'}(\boldsymbol{\xi}) = \langle \partial_\alpha \psi_j(\boldsymbol{\xi}) | \partial_\beta \psi_{j'}(\boldsymbol{\xi}) \rangle - \langle \partial_\alpha \psi_j(\boldsymbol{\xi}) | P(\boldsymbol{\xi}) | \partial_\beta \psi_{j'}(\boldsymbol{\xi}) \rangle, \tag{3.38}$$

where now $P(\boldsymbol{\xi})$ is the single-particle projector; even this matrix is gauge-covariant. If we rewrite the Cartesian components of Eq. (3.35) as $\mathscr{A}_{\alpha,jj'}(\boldsymbol{\xi}) = -\text{Im } \langle \psi_j(\boldsymbol{\xi}) | \partial_\alpha \psi_{j'}(\boldsymbol{\xi}) \rangle$, the nonAbelian curvature becomes

$$\Omega_{\alpha\beta,jj'}(\boldsymbol{\xi}) = \partial_\alpha \mathscr{A}_{\beta,jj'}(\boldsymbol{\xi}) - \partial_\beta \mathscr{A}_{\alpha,jj'}(\boldsymbol{\xi}) - i[\mathscr{A}_\alpha(\boldsymbol{\xi}), \mathscr{A}_\beta(\boldsymbol{\xi})]_{jj'}. \tag{3.39}$$

With respect to Eq. (3.15) notice the presence of an extra term, which vanishes in the Abelian case. At fixed jj' Eq. (3.39) is clearly antisymmetric in the $\alpha\beta$ Cartesian indices, while at fixed $\alpha\beta$ it is an Hermitian $n \times n$ matrix.

The trace over the j index of the (nonAbelian) metric-curvature tensor equals the corresponding (Abelian) metric-curvature tensor of the many-body ground state, Eq. (3.25):

$$\begin{aligned} & \sum_{j=1}^{n} \mathscr{F}_{\alpha\beta,jj}(\boldsymbol{\xi}) \\ = \; & \sum_{j=1}^{n} \langle \partial_\alpha \psi_j(\boldsymbol{\xi}) | \partial_\beta \psi_j(\boldsymbol{\xi}) \rangle - \sum_{jj'=1}^{n} \langle \partial_\alpha \psi_j(\boldsymbol{\xi}) | \psi_{j'}(\boldsymbol{\xi}) \rangle \langle \psi_{j'}(\boldsymbol{\xi}) | \partial_\beta \psi_j(\boldsymbol{\xi}) \rangle \\ = \; & \mathcal{F}_{\alpha\beta}(\boldsymbol{\xi}). \end{aligned} \tag{3.40}$$

The proof of the last line of Eq. (3.40) is tedious, but straightforward.

3.9 Bloch orbitals

We have remained very general so far. The case where the parameter $\boldsymbol{\xi}$ coincides with the Bloch vector $\mathbf{k}$ bears a particular relevance in the context of the present Notes. In the framework of first-principle calculations for crystalline systems, the Bloch states $|\psi_{j\mathbf{k}}\rangle$ are the KS orbitals.

The domain where the $\mathbf{k}$ parameter varies (the reciprocal cell or, equivalently, the Brillouin zone, BZ) has the geometry of a torus in 1d, 2d, 3d. The whole BZ is a closed surface; we denote as $\mathbf{G}$ a reciprocal vector.

The Bloch orbital of the j-th band is $|\psi_{j\mathbf{k}}\rangle = \mathrm{e}^{i\mathbf{k}\cdot\mathbf{r}}|u_{j\mathbf{k}}\rangle$, where $|u_{j\mathbf{k}}\rangle$ are BZ-periodical, and are eigenfunctions of the Hamiltonian $H_{\mathbf{k}} = \mathrm{e}^{-i\mathbf{k}\cdot\mathbf{r}}H\mathrm{e}^{i\mathbf{k}\cdot\mathbf{r}}$. While the $|\psi_{j\mathbf{k}}\rangle$ at different $\mathbf{k}$'s are mutually orthogonal, the $|u_{j\mathbf{k}}\rangle$ live instead in the same Hilbert space (they are all BZ-periodical).

The physical meaning of all the mathematical quantities introduced next will be discussed in the following Chapters, and not anticipated in the present one.

The Berry connection of the j-th orbital is

$$\boldsymbol{\mathcal{A}}_j(\mathbf{k}) = i\langle u_{j\mathbf{k}}|\nabla_{\mathbf{k}} u_{j\mathbf{k}}\rangle. \tag{3.41}$$

The relative phases at different $\mathbf{k}$'s are arbitrary. Whenever possible, it is customary to enforce the so-called periodic gauge $|\psi_{j\mathbf{k}+\mathbf{G}}\rangle = |\psi_{j\mathbf{k}}\rangle$, which implies

$$|u_{j\mathbf{k}+\mathbf{G}}\rangle = \mathrm{e}^{-i\mathbf{G}\cdot\mathbf{r}}|u_{j\mathbf{k}}\rangle. \tag{3.42}$$

We stress, however, that in topologically nontrivial crystals it is generally *impossible* to adopt a periodic gauge.

The interesting closed paths C on the torus are lines across the reciprocal cell, from one face to the opposite one. For an insulator with n occupied bands the Berry phase is, according to the previous section,

$$\gamma = i\sum_{j=1}^{n}\int_C \boldsymbol{\mathcal{A}}_j(\mathbf{k})\cdot d\mathbf{k} = i\sum_{j=1}^{n}\int_C d\mathbf{k}\cdot\langle u_{j\mathbf{k}}|\nabla_{\mathbf{k}} u_{j\mathbf{k}}\rangle. \tag{3.43}$$

This Berry phase depends on the choice of the origin in the crystal cell. For centrosymmetric crystals, if the origin is at a center of inversion symmetry, the only allowed values are $\gamma = 0$ and $\gamma = \pi$ (modulo 2π).

In case of band crossings the definition of individual bands is ambiguous, but the Berry phase in Eq. (3.43) is not affected by such ambiguity. More generally, the results of the previous Section show that Eq. (3.43) is invariant by a nonAbelian gauge transformation, i.e. by mixing of the occupied orbitals between themselves by an arbitrary unitary matrix $U_{jj'}(\mathbf{k})$. The mixed orbitals are no longer Hamiltonian eigenstates; any gauge where instead the $|u_{j\mathbf{k}}\rangle$ are eigenstates will be called "Hamiltonian gauge".

The discretization of Eq. (3.43) is nowadays implemented in most electronic structure codes [56, 57, 58, 59] in order to compute the macroscopic polarization of crystalline dielectrics (Sec. 5.3). Such discretization is based on the following result: if $|\Psi(\mathbf{k})\rangle$ is the Slater determinant of the n occupied $|u_{j\mathbf{k}}\rangle$, then

$$\langle\Psi(\mathbf{k}_1)|\Psi(\mathbf{k}_2)\rangle = \det\, S(\mathbf{k}_1,\mathbf{k}_2), \tag{3.44}$$

where $S(\mathbf{k}_1,\mathbf{k}_2)$ is the $n\times n$ overlap matrix

$$S_{jj'}(\mathbf{k}_1,\mathbf{k}_2) = \langle u_{j\mathbf{k}_1}|u_{j'\mathbf{k}_2}\rangle. \tag{3.45}$$

We discretize with $M{+}1$ points on the line, where $\mathbf{k}_{M+1} = \mathbf{k}_1+\mathbf{G}$; it is understood that $|u_{j\mathbf{k}_{M+1}}\rangle = \mathrm{e}^{-i\mathbf{G}\cdot\mathbf{r}}|u_{j\mathbf{k}_1}\rangle$. The discretized formula is then

$$\begin{aligned}\gamma &= i\int_C d\mathbf{k}\cdot\langle\Psi(\mathbf{k})|\nabla_{\mathbf{k}}\Psi(\mathbf{k})\rangle \rightarrow -\mathrm{Im}\,\log\,\Pi_{s=1}^{M}\langle\Psi(\mathbf{k}_s)|\Psi(\mathbf{k}_{s+1})\rangle \\ &= -\mathrm{Im}\,\log\,\det\,\Pi_{s=1}^{M} S(\mathbf{k}_s,\mathbf{k}_{s+1}).\end{aligned} \tag{3.46}$$

Notice that this is *numerically* gauge invariant, i.e. it does not depend on how the diagonalization routine chooses the phases and/or the ordering of the eigenvectors.

The metric-curvature tensor of n occupied bands obtains straightforwardly from Eq. (3.40):

$$\mathcal{F}_{\alpha\beta}(\mathbf{k})=\sum_{j=1}^{n}\langle\partial_\alpha u_{j\mathbf{k}}|\partial_\beta u_{j\mathbf{k}}\rangle-\sum_{jj'=1}^{n}\langle\partial_\alpha u_{j\mathbf{k}}|u_{j'\mathbf{k}}\rangle\langle u_{j'\mathbf{k}}|\partial_\beta u_{j\mathbf{k}}\rangle. \tag{3.47}$$

This is usually integrated over the BZ; being gauge invariant, it carries in principle physical meaning even *before* any integration. Indeed, the **k**-dependent (single-band) curvature enters the theory of semiclassical transport in crystalline solids [60, 61, 12], Sec. 4.5.

Eq. (3.47) is the trace of the nonAbelian metric-curvature whose expression is

$$\begin{aligned}\mathscr{F}_{\alpha\beta,jj'}(\mathbf{k}) &= \langle\partial_\alpha u_{j\mathbf{k}}|\partial_\beta u_{j'\mathbf{k}}\rangle-\sum_{j"=1}^{n}\langle\partial_\alpha u_{j\mathbf{k}}|u_{j"\mathbf{k}}\rangle\langle u_{j"\mathbf{k}}|\partial_\beta u_{j'\mathbf{k}}\rangle\\ &= \sum_{s=n+1}^{\infty}\langle\partial_\alpha u_{jk}|u_{s\mathbf{k}}\rangle\langle u_{s\mathbf{k}}|\partial_\beta u_{j'\mathbf{k}}\rangle.\end{aligned} \tag{3.48}$$

The nonAbelian metric and curvature are

$$\begin{aligned}g_{\alpha\beta,jj'} &= \frac{1}{2}[\,\mathscr{F}_{\alpha\beta,jj'}(\mathbf{k})+\mathscr{F}_{\beta\alpha,jj'}(\mathbf{k})\,]\\ \Omega_{\alpha\beta,jj'} &= i[\,\mathscr{F}_{\alpha\beta,jj'}(\mathbf{k})-\mathscr{F}_{\beta\alpha,jj'}(\mathbf{k})\,].\end{aligned} \tag{3.49}$$

The many-band curvature obtains from either Eq. (3.47) or Eq. (3.49) as

$$\boldsymbol{\Omega}_{\alpha\beta}=-2\,\mathrm{Im}\,\sum_{j=1}^{n}\langle\partial_\alpha u_{j\mathbf{k}}|\partial_\beta u_{j\mathbf{k}}\rangle. \tag{3.50}$$

Chapter 4

Manifestations of the Berry phase

4.1 A toy-model Hamiltonian

Our toy model here is a two-level spinless Hamiltonian, of the form

$$\begin{aligned} H(\boldsymbol{\xi}) &= \boldsymbol{\xi} \cdot \vec{\sigma} \\ &= \xi\, (\sin\vartheta \cos\varphi \;\sigma_x + \sin\vartheta \sin\varphi \;\sigma_y + \cos\vartheta \;\sigma_z), \end{aligned} \tag{4.1}$$

where σ_α are the three Pauli matrices. The spectrum is non degenerate for $\boldsymbol{\xi} \neq 0$, and the lowest eigenvalue is $-\xi$. Upon symmetry arguments, we can already guess the curvature to be isotropic.

4.1.1 Connection and curvature

The lowest eigenvector is

$$|\psi(\vartheta, \varphi)\rangle = \begin{pmatrix} \sin\frac{\vartheta}{2} \mathrm{e}^{-i\varphi} \\ -\cos\frac{\vartheta}{2} \end{pmatrix}.$$

This corresponds to a specific gauge choice; the eigenvector can be multiplied by an overall (ϑ, φ)-dependent phase factor. The Berry connection and curvature are

$$\begin{aligned} \mathcal{A}_\vartheta &= i\langle\psi|\partial_\vartheta \psi\rangle = 0 \\ \mathcal{A}_\varphi &= i\langle\psi|\partial_\varphi \psi\rangle = \sin^2\frac{\vartheta}{2} \\ \boldsymbol{\Omega} &= \partial_\vartheta \mathcal{A}_\varphi - \partial_\varphi \mathcal{A}_\vartheta = \frac{1}{2}\sin\vartheta. \end{aligned} \tag{4.2}$$

The curvature is gauge-invariant, while the connection is gauge-dependent. Within our gauge choice the connection displays a vortex at the south pole ($\vartheta = \pi$); other gauges yield the singularity at a different point, but a singularity is unavoidable. It is impossible to find a gauge which is smooth on the whole closed surface, and a nonsingular connection; the singularity is often called an "obstruction". The algebra is the same as for Dirac's theory of the magnetic monopole [44, 46]: the degeneracy at the origin is the monopole, and the singularity is the "Dirac string".

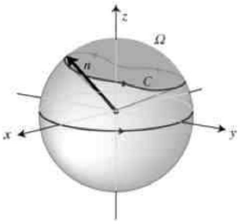

Figure 4.1: A closed curve C on the surface of the sphere, and the solid angle spanned by it.

4.1.2 Chern number

The domain of the parameters (ϑ, φ) is a rectangle, which indeed has the topology of a torus: a closed surface. Integrating the Berry curvature therein we get

$$\frac{1}{2\pi} \int_{S^2} d\vartheta d\varphi \, \frac{1}{2} \sin \vartheta = 1, \tag{4.3}$$

i.e. the Chern number of the lowest eigenstate in this problem. This integer measures the strength of the singularity (magnetic monopole), which resides in a site *inaccessible* to the quantum system. The highest eigenstate has $C_1 = -1$, since the total Chern number is zero.

This simple example illustrates well the meaning of a topological invariant of the quantum mechanical ground state. The Chern number is in fact very robust under continuous deformations of the surface and of the Hamiltonian: its value is always one insofar as one (and only one) degeneracy point is included in the closed surface.

4.1.3 Berry phase

Suppose now we evaluate the Berry phase over any closed curve C on the sphere (Fig. 4.1)

$$\gamma = \oint_C \boldsymbol{\mathcal{A}}(\boldsymbol{\xi}) \cdot d\boldsymbol{\xi}. \tag{4.4}$$

Owing to Stokes theorem, the Berry phase for this toy model problem clearly equals the solid angle spanned by the curve, divided by 2. The inherent 4π arbitrariness in the solid angle leads to the well known 2π arbitrariness in the Berry phase: for instance at the equator $\gamma = \pi$ modulo 2π. If we cut the sphere in two hemispheres, as in Sec. 3.4 (see Fig. 3.4), the difference of the boundary Berry phases equals 0 modulo 2π, i.e. 2π times an integer, as it must be. But in order to tell *which* integer (the actual Chern number) the two connections are useless: one has to integrate the curvature, as in Eq. (4.3). Despite this feature, in numerical work Chern numbers are typically computed via Berry phases, as described in the next Section.

4.1.4 Numerical considerations

This exactly soluble example also provides the occasion for illustrating the standard computational approach to Chern numbers. Suppose we discretize the (ϑ, φ) domain with a rectangular mesh, and that we diagonalize the Hamiltonian at the points of the mesh. The gauge at any point is chosen by the

diagonalization routine and is thus erratic; we only enforce the toroidal topology by requiring that the phases at the opposite edges of the rectangle are the same.

Then for each small rectangle we compute the discrete Berry phase as in Eq. (3.7), i.e.

$$\begin{aligned}\gamma &= -\operatorname{Im}\log \langle\psi(\vartheta,\varphi)|\psi(\vartheta+\Delta\vartheta,\varphi)\rangle\langle\psi(\vartheta+\Delta\vartheta,\varphi)|\psi(\vartheta+\Delta\vartheta,\varphi+\Delta\varphi)\rangle \\ &\times \langle\psi(\vartheta+\Delta\vartheta,\varphi+\Delta\varphi)|\psi(\vartheta,\varphi+\Delta\varphi)\rangle\langle\psi(\vartheta,\varphi+\Delta\varphi)|\psi(\vartheta,\varphi)\rangle. \end{aligned} \tag{4.5}$$

The Berry curvature is the Berry phase per unit (ϑ,φ) area. In this simple, analytically soluble, case we know the exact value; Eq. (4.2) implies for Eq. (4.5) $\gamma = \frac{1}{2}\sin\vartheta\,\Delta\vartheta\Delta\varphi$ modulo 2π. The Chern number is the integral over the domain, and is therefore equal to the sum of all the phases computed as in Eq. (4.5) and covering the whole domain.

Last but not least: how do we get rid of the modulo 2π indeterminacy in Eq. (4.5)? The size of $\Delta\vartheta\Delta\varphi$ is very small, and each contribution γ to the sum is also small (proportional to $\Delta\vartheta\Delta\varphi$), although Eq. (4.5) is in principle arbitrary modulo 2π. It should be now pretty clear that the right solution is in choosing the Im log branch with values in $[-\pi,\pi]$.

4.2 Early discoveries reinterpreted

4.2.1 Aharonov-Bohm effect

Here we reformulate the Aharonov-Bohm effect as a special case of a Berry phase. Suppose we have an electron in a box (infinite potential well) centered at the origin. We take the ground wavefunction as real, and we write it as $\chi(\mathbf{r})$. The time-independent Schrödinger equation is:

$$\left[\frac{p^2}{2m} + V(\mathbf{r})\right]\chi(\mathbf{r}) = E\chi(\mathbf{r}). \tag{4.6}$$

Displacing the center of the box at position $\mathbf{R}$ changes the Hamiltonian to

$$H(\mathbf{R}) = \frac{p^2}{2m} + V(\mathbf{r}-\mathbf{R}) \ : \tag{4.7}$$

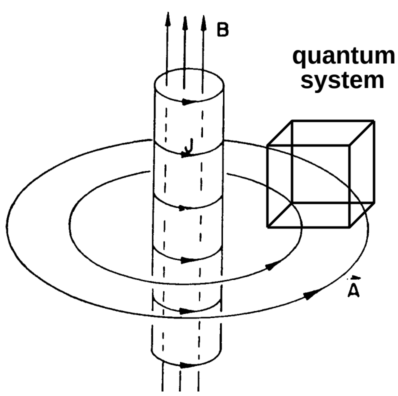

Figure 4.2: A particle in a box, transported round a solenoid

we will identify the $\boldsymbol{\xi}$ parameter with the box position $\mathbf{R}$. Because of translational invariance, the $\mathbf{R}$-dependence of the state vectors is

$$\langle \mathbf{r}|\psi(\mathbf{R})\rangle = \chi(\mathbf{r}-\mathbf{R}), \tag{4.8}$$

while the eigenvalue is $\mathbf{R}$-independent.

Suppose now that a magnetic field is switched on somewhere in space. Then the Hamiltonian becomes

$$H(\mathbf{R}) = \frac{1}{2m}\left[\mathbf{p}+\frac{e}{c}\mathbf{A}(\mathbf{r})\right]^2 + V(\mathbf{r}-\mathbf{R}) \,, \tag{4.9}$$

where $\mathbf{A}$ is the vector potential and e is the electron charge. It can be easily verified that a solution of the Schrödinger equation can be formally written in the form:

$$\langle \mathbf{r}|\psi(\mathbf{R})\rangle = \exp\left(-\frac{ie}{\hbar c}\int_{\mathbf{R}}^{\mathbf{r}} \mathbf{A}(\mathbf{r}')\cdot d\mathbf{r}'\right)\chi(\mathbf{r}-\mathbf{R}). \tag{4.10}$$

But such a solution in general is *not* a single-valued function of $\mathbf{r}$, since the phase factor depends on the path. Therefore we restrict ourselves to a less general case, where the magnetic field is generated by a solenoid: the $\mathbf{B}$ field is nonzero only within a given cylinder, and we don't allow our box to overlap this cylinder by suitably restricting the domain of $\mathbf{R}$. This situation is sketched in Fig. 4.2. With such a choice the wavefunction, Eq. (4.10), is a single valued function of $\mathbf{r}$ for any fixed $\mathbf{R}$, and is therefore an honest ground wavefunction. As for the dependence on $\mathbf{R}$, Eq. (4.10) only guarantees *local* single-valuedness, since the domain is not simply connected: when the system is transported on a closed path winding once round the solenoid, the electron wavefunction picks up a Berry's phase. This phase difference can be actually detected in interference experiments.

The Berry connection of the problem is

$$\boldsymbol{\mathcal{A}}(\mathbf{R}) = i\langle\psi(\mathbf{R})|\nabla_{\mathbf{R}}\psi(\mathbf{R})\rangle = -\frac{e}{\hbar c}\mathbf{A}(\mathbf{R}) - i\int d\mathbf{r}\;\chi(\mathbf{r}-\mathbf{R})\nabla_{\mathbf{R}}\chi(\mathbf{r}-\mathbf{R}), \tag{4.11}$$

where the last term vanishes. Therefore in the present case the Berry connection is proportional to the ordinary vector potential. A gauge transformation in the quantum-mechanical sense also coincides with an electromagnetic gauge transformation, which changes $\mathbf{A}$ while leaving $\mathbf{B}$ invariant. In fact, in this example $\mathbf{B}$ is essentially the Berry curvature. The Berry phase is

$$\gamma = -\frac{e}{\hbar c}\oint_C \mathbf{A}(\mathbf{R})\cdot d\mathbf{R} = -\frac{2\pi}{\phi_0}\oint_C \mathbf{A}(\mathbf{R})\cdot d\mathbf{R}, \tag{4.12}$$

where ϕ_0 is the flux quantum. Therefore γ measures the flux of the magnetic field across the *interior* of the solenoid, a space region *not accessed* by the quantum system: above, we have called it "inaccessible flux". Only the fractional part of the flux has physical meaning.

4.2.2 Molecular Aharonov-Bohm effect

Here we identify the "slow coordinate" $\boldsymbol{\xi}$ with a d-dimensional nuclear coordinate, and the state vector $|\Psi(\boldsymbol{\xi})\rangle$ with the electronic ground-state wavefunction in the Born-Oppenheimer approximation.

We start from the complete Hamiltonian $\mathcal{H}$ of an isolated molecular system, and we explicitly separate the nuclear kinetic energy:

$$\mathcal{H}(\boldsymbol{\xi}, [\mathbf{x}]) = \frac{1}{2} \sum_{\alpha,\beta=1}^{d} \mathfrak{M}^{-1}_{\alpha\beta}\, p_\alpha p_\beta + H(\boldsymbol{\xi}, [\mathbf{x}]), \tag{4.13}$$

where $[\mathbf{x}]$ indicates the electronic degrees of freedom collectively, $p_\alpha = -i\hbar\, \partial/\partial\xi_\alpha$ is the canonical momentum conjugated to ξ_α, and the inverse mass matrix $\mathfrak{M}^{-1}$ in general may be a function of $\boldsymbol{\xi}$, but not of the momenta.

The Born-Oppenheimer approximation starts by writing the eigenfunctions of Eq. (4.13) in the Schrödinger representation as the product $\langle\, [\mathbf{x}]\, |\Psi(\boldsymbol{\xi})\rangle\, \Phi(\boldsymbol{\xi})$. Our aim is obtaining an effective Schrödinger equation for the nuclear wavefunction $\Phi(\boldsymbol{\xi})$, where the electronic degrees of freedom have been integrated out. We start considering the effect of the canonical nuclear momentum $\mathbf{p}$ on the product ansatz:

$$\mathbf{p}\, |\Psi(\boldsymbol{\xi})\rangle\, \Phi(\boldsymbol{\xi}) = -i\hbar\, |\Psi(\boldsymbol{\xi})\rangle\, \nabla_{\boldsymbol{\xi}} \Phi(\boldsymbol{\xi}) - i\hbar\, |\nabla_{\boldsymbol{\xi}} \Psi(\boldsymbol{\xi})\rangle\, \Phi(\boldsymbol{\xi}). \tag{4.14}$$

We then multiply by the electronic eigenbra $\langle\Psi(\boldsymbol{\xi})|$ on the left, thus integrating over the electronic degrees of freedom. We get the effective nuclear momentum $\boldsymbol{\pi}$ acting on Φ as:

$$\boldsymbol{\pi}\ \Phi(\boldsymbol{\xi}) = [\ \mathbf{p} - i\hbar\langle\Psi(\boldsymbol{\xi})|\nabla_{\boldsymbol{\xi}}\Psi(\boldsymbol{\xi})\rangle\ ]\ \Phi(\boldsymbol{\xi}) = [\ \mathbf{p} - \hbar\boldsymbol{\mathcal{A}}(\xi)\ ]\ \Phi(\boldsymbol{\xi}), \tag{4.15}$$

where we easily recognize the Berry connection. The momentum $\boldsymbol{\pi}$ is the kinematical (also called covariant, or mechanical) momentum, to be distinguished from $\mathbf{p} = -i\hbar\nabla_{\boldsymbol{\xi}}$, which is the canonical momentum.

Whenever the time scales of nuclear and electronic motions are well separated the coupling between different electronic states can be neglected, and the adiabatic approximations allows to treat the slow variable $\boldsymbol{\xi}$ in $H(\boldsymbol{\xi}, [\mathbf{x}])$ as a classical parameter. The electronic eigenvalue $E(\boldsymbol{\xi})$ of a given state (e.g. the ground state) plays therefore the role of a (scalar) potential for nuclear motion, whose effective Hamiltonian acting on $\Phi(\boldsymbol{\xi})$ is then:

$$\mathsf{H}_{\text{eff}} = \frac{1}{2} \sum_{\alpha,\beta=1}^{d} \mathfrak{M}^{-1}_{\alpha\beta}\, \pi_\alpha \pi_\beta + E(\boldsymbol{\xi}). \tag{4.16}$$

In the molecular physics literature the extra term in Eq. (4.15) is seldom mentioned, and $\boldsymbol{\pi}$ is identified with $\mathbf{p}$. The reason is that for a time-reversal-invariant Hamiltonian, and in absence of spin-orbit interaction, the wave function can always be taken as real. This corresponds to the parallel transport gauge, and the Berry connection vanishes at all $\boldsymbol{\xi}$; the tradeoff is that—in some cases—the electronic wave function is *not* single valued along a closed path: see Fig. 2.3. The alternative approach, due to Mead and Truhlar [30, 62], is to choose a different gauge, where the electronic wave function is single valued and complex. The Berry phase is gauge invariant; the values allowed by time-reversal symmetry are 0 and π; the two cases are experimentally distinguishable.

We stress that, whenever the ionic motion is purely classical and governed by Newton's equation, the vector-potential-like term in Eq. (4.15) is irrelevant: the corresponding curvature (magnetic-field-like) is in fact identically vanishing

along the nuclear trajectory on the Born-Oppenheimer surface. We anticipate that the case where a genuine magnetic field is present—and the Hamiltonian is no longer time-reversal-invariant—is *qualitatively* different in this respect, see Sec. 4.3 below.

4.2.3 Integer quantum Hall effect

The famous TKNN (Thouless, Kohmoto, Nightingale, and den Nijs) paper appeared in 1982 [47] and marks the very first occurrence of a topological invariant (the first Chern number) in electronic structure. The outline provided here is inspired by Kohmoto [63].

We consider a two-dimensional independent-electron system in a lattice-periodical potential, and subject to a perpendicular $\mathbf{B}$ field. The Hamiltonian is *not* translationally invariant, but one can address the magnetic translation group. We choose a large enough "supercell", such that the magnetic flux is commensurate (i.e. an integer number of flux quanta ϕ_0 thread the supercell): in this case a continuous $\mathbf{k}$ vector can be defined in the magnetic Brillouin zone.

As in Sec. 3.9,we define $|\psi_{j\mathbf{k}}\rangle = \mathrm{e}^{i\mathbf{k}\cdot\mathbf{r}}|u_{j\mathbf{k}}\rangle$ and $H_{\mathbf{k}} = \mathrm{e}^{-i\mathbf{k}\cdot\mathbf{r}}H\mathrm{e}^{i\mathbf{k}\cdot\mathbf{r}}$; the latter takes here the form

$$H_{\mathbf{k}} = \frac{1}{m}\left[\mathbf{p} + \hbar\mathbf{k} + \frac{e}{c}\mathbf{A}(\mathbf{r})\right]^2 + \mathscr{V}(\mathbf{r}), \tag{4.17}$$

where $\mathscr{V}$ is the substrate potential. The velocity can be expressed as

$$\mathbf{v} = \frac{1}{\hbar}\nabla_{\mathbf{k}}H_{\mathbf{k}}, \tag{4.18}$$

a formula often recurring in the present Notes in various forms, see e.g. Eqs. (1.18) and (2.17).

The Kubo formula for transverse conductivity is [64]

$$\begin{aligned}
\sigma_{xy} &= 2\hbar e^2 \,\mathrm{Im} \sum_{jj'} \frac{1}{(2\pi)^2}\int_{\epsilon_{j\mathbf{k}}<\epsilon_{\mathrm{F}}<\epsilon_{j'\mathbf{k}}} d\mathbf{k}\, \frac{\langle u_{j\mathbf{k}}|v_x|u_{j'\mathbf{k}}\rangle\langle u_{j'\mathbf{k}}|v_x|u_{j\mathbf{k}}\rangle}{(\epsilon_{j\mathbf{k}} - \epsilon_{j'\mathbf{k}})^2} \qquad (4.19)\\
&= 2\frac{e^2}{\hbar}\,\mathrm{Im} \sum_{jj'} \frac{1}{(2\pi)^2}\int_{\epsilon_{j\mathbf{k}}<\epsilon_{\mathrm{F}}<\epsilon_{j'\mathbf{k}}} d\mathbf{k}\, \frac{\langle u_{j\mathbf{k}}|\partial_x H_{\mathbf{k}}|u_{j'\mathbf{k}}\rangle\langle u_{j'\mathbf{k}}|\partial_y H_{\mathbf{k}}|u_{j\mathbf{k}}\rangle}{(\epsilon_{j\mathbf{k}} - \epsilon_{j'\mathbf{k}})^2}.
\end{aligned}$$

If we now consider the case where the Fermi level lies in a gap, with n filled bands (Landau levels in a flat potential), Eq. (4.19) becomes the BZ integral

$$\sigma_{xy} = 2\frac{e^2}{\hbar}\,\mathrm{Im}\,\frac{1}{(2\pi)^2}\int_{\mathrm{BZ}} d\mathbf{k} \sum_{j=1}^{n}\sum_{j'=n+1}^{\infty} \frac{\langle u_{j\mathbf{k}}|\partial_x H_{\mathbf{k}}|u_{j'\mathbf{k}}\rangle\langle u_{j'\mathbf{k}}|\partial_y H_{\mathbf{k}}|u_{j\mathbf{k}}\rangle}{(\epsilon_{j\mathbf{k}} - \epsilon_{j'\mathbf{k}})^2}. \tag{4.20}$$

The integrand is just a simple generalization of the sum-over-states formula of Eq. (3.28). Using the same arguments as in Ch. 3 it is rather straightforward to arrive at the identity

$$\begin{aligned}
&\mathrm{Im}\sum_{j=1}^{n}\sum_{j'=n+1}^{\infty} \frac{\langle u_{j\mathbf{k}}|\partial_x H_{\mathbf{k}}|u_{j'\mathbf{k}}\rangle\langle u_{j'\mathbf{k}}|\partial_y H_{\mathbf{k}}|u_{j\mathbf{k}}\rangle}{(\epsilon_{j\mathbf{k}} - \epsilon_{j'\mathbf{k}})^2}\\
= \;&\mathrm{Im}\sum_{j=1}^{n}\langle\partial_x u_{j\mathbf{k}}|\partial_x u_{j\mathbf{k}}\rangle = -\frac{1}{2}\boldsymbol{\Omega}_{xy}(\mathbf{k}), \qquad (4.21)
\end{aligned}$$

where the many-band Berry curvature, Eq. (3.50), appears. Since the BZ is a torus, the BZ integral of the curvature equals 2π times an integer, the (first) Chern number C. The milestone TKNN discovery is that Hall conductivity is a Chern number when expressed in klitzing^{-1}:

$$\sigma_{xy} = -\frac{e^2}{h}C. \tag{4.22}$$

Notice that the sign choices are not uniform across the literature.

Conductivity is a property of the *excitations* of the system, as it is perspicuous in the Kubo formula above. The Chern number, instead, is a *ground state* property. The identity relating them belongs to the general class of fluctuation-dissipation theorems, although this looks like an oxymoron, the Hall conductivity being here dissipationless. The interpretation of the Chern number as a ground-state quantum fluctuation will be elaborated somewhere else in the present Notes.

The topological nature of the observable explains its extreme robustness under variations of magnetic field, carrier density, substrate disorder, and more. The topological invariant C is identified with the filling ν using the same arguments as in Sec. 2.4.4; the integer can only be varied by crosssing a conducting state.

4.3 Adiabatic approximation in a magnetic field

The general problem of the nuclear motion—both classical and quantum—in presence of an external magnetic field has been first solved in 1988 by Schmelcher, Cederbaum, and Meyer [65]. It is remarkable that such a fundamental problem was solved so late, and that even today the relevant literature is ignored by textbooks and little cited. The solution is a manifestation of geometrical effects in electronic wavefunctions [66], which appears in a spectacular way even when the nuclear motion is addressed at the classical level.

When a genuine magnetic field, generated by some external source, acts on the molecular system, the Hamiltonian of Eq. (4.13) is modified by the addition of a vector potential term in the kinetic energies of both the nuclei and the electrons. Proceeding as in the zero field case, one writes an ansatz wavefunction and arrives at the effective Hamiltonian for the ionic motion, Eq. (4.16), where an extra term must be added to the kinematical momentum $\boldsymbol{\pi}$ of Eq. (4.15). There are thus *two* vector potentials in the effective nuclear Hamiltonian: a geometric one, and a genuinely magnetic one.

However, with respect to the zero-field case, there is a qualitative difference whose importance is overwhelming. Since the electronic Hamiltonian is no longer invariant under time-reversal, the electronic wavefunction is necessarily complex, and the curvature is in general nonzero. No singularity is needed to produce geometrical effects on the nuclear motion; the Berry phase will be in general nonzero on any path in the space of nuclear coordinates.

Suppose we are interested into the nuclear motion at the purely classical level. The Hamiltonian of Eq. (4.16)—whose kinematical momentum $\boldsymbol{\pi}$ includes now the two different vector potentials—yields the Hamilton equations of motion, which can be transformed into the Newton equations of motion: within the latter, the effects of the vector potentials appear in terms of fields, in the form

of Lorentz forces. The curl of the magnetic vector potential obviously yields the magnetic field due to the external source; the curl of the geometric vector potential (Berry curvature) yields an additional "magnetic-like" field which is *nonzero* even on the classical trajectory of the nuclei. We stress that this is at variance with the zero-field case, where the Berry phase had no effect on the ionic motion at the classical level, and could only be detected when quantizing the ionic degrees of freedom.

Within a naïve Born–Oppenheimer approximation—where Berry phases are neglected—the magnetic field acts on the nuclei as it they were "naked" charges: a proper treatment must instead account for electronic screening: this is provided by the geometric vector potential. Surprisingly, there are very few calculations of the effect: it is pretty clear, however, that the geometric term is no small correction.

For pedagogical purposes we consider the case of a hydrogen atom, hence we identify the electronic degrees of freedom $[\mathbf{x}]$, used in Sec. 4.2.2, with a single coordinate $\mathbf{r}$, and the parameter $\boldsymbol{\xi}$ with the nuclear coordinate $\mathbf{R}$. If the atom is subject only to a magnetic field, the complete Hamiltonian $\mathcal{H}$ and the electronic Hamiltonian H are

$$\mathcal{H}(\mathbf{R},\mathbf{r}) = \frac{1}{2M}\left[\mathbf{p} - \frac{e}{c}\mathbf{A}(\mathbf{R})\right]^2 + H(\mathbf{R},\mathbf{r}); \tag{4.23}$$

$$H(\mathbf{R},\mathbf{r}) = \frac{1}{2m}\left[-i\hbar\nabla_{\mathbf{r}} + \frac{e}{c}\mathbf{A}(\mathbf{r})\right]^2 - \frac{e^2}{|\mathbf{r}-\mathbf{R}|}. \tag{4.24}$$

As explained above, the nuclear kinematical momentum of Eq. (4.15) becomes

$$\boldsymbol{\pi} = \mathbf{p} - \hbar\boldsymbol{\mathcal{A}}(\mathbf{R}) - \frac{e}{c}\mathbf{A}(\mathbf{R}) \tag{4.25}$$

The case of a constant $\mathbf{B}$ field can be dealt with analytically. We choose the central gauge $\mathbf{A}(\mathbf{r}) = \frac{1}{2}\mathbf{B}\times\mathbf{r}$. If $\phi(\mathbf{r})$ is the exact ground eigenfunction when the proton sits at $\mathbf{R} = 0$, the eigenfunction at a generic $\mathbf{R}$ is:

$$\langle\mathbf{r}|\psi(\mathbf{R})\rangle = \mathrm{e}^{-\frac{ie}{2\hbar c}\mathbf{r}\cdot\mathbf{B}\times\mathbf{R}}\phi(\mathbf{r}-\mathbf{R}), \tag{4.26}$$

with an $\mathbf{R}$-independent eigenvalue. The Berry connection is clearly

$$\begin{aligned}\boldsymbol{\mathcal{A}}(\mathbf{R}) &= i\langle\psi(\mathbf{R})|\nabla_{\mathbf{R}}\psi(\mathbf{R})\rangle = -\frac{e}{2\hbar c}\langle\psi(\mathbf{R})|\mathbf{B}\times\mathbf{r}|\psi(\mathbf{R})\rangle = -\frac{e}{2\hbar c}\mathbf{B}\times\mathbf{R} \\ &= -\frac{e}{\hbar c}\mathbf{A}(\mathbf{R}),\end{aligned} \tag{4.27}$$

since the $\mathbf{R}$-derivative of $\phi(\mathbf{r}-\mathbf{R})$ does not contribute. Replacing Eq. (4.27) into Eq. (4.25) we find $\boldsymbol{\pi} = \mathbf{p}$, as it must be: the nucleus travels at constant speed, and is not deflected by a Lorentz force.

Remarkably, the "magnetic-like" field due to the Berry phase is—in this simple example—exactly opposite to the external magnetic field, thus providing the complete screening which is physically expected. In less trivial situations, the screening affects significantly the molecular vibrations and the classical nuclear motion in general. The case of H_2 has been investigated in 2007 by Ceresoli et al. [67].

4.4 Anomalous Hall effect

Edwin H. Hall discovered the eponymous effect, nowadays in all solid state textbooks, in 1879. Shortly later, in 1881, he discovered that the effect in ferromagnetic materials is "anomalous". While in nonmagnetic materials the Hall resistivity is linear in the magnetic field, in ferromagnetic ones it saturates to a value roughly proportional to the macroscopic magnetization.

In 1954 Karplus and Luttinger [68] provided a theory of the anomalous Hall effect (AHE) which, with hindsight, was indeed geometrical. The understanding of AHE, however, remained controversial for another 40 years and more. One of the reasons for this state of affairs is that the intrinsic geometrical contribution is partly obscured by extrinsic contributions (jargon: skew scattering and side jumps), not easily disentangled in the experimental data. A comprehensive review appeared very recently [11]. Here we only outline the intrinsic theory, based on geometrical concepts, which owes to a couple of papers appeared in 2002, by Jungwith, Niu, and MacDonald [69], and by Onoda and Nagaosa [70].

The starting point is the Kubo formula of Eq. (4.19), which we rewrite in dimension d as

$$\sigma_{xy} = 2\hbar e^2 \,\mathrm{Im} \sum_{jj'} \frac{1}{(2\pi)^d} \int_{\epsilon_{j\mathbf{k}}<\epsilon_\mathrm{F}<\epsilon_{j'\mathbf{k}}} d\mathbf{k}\, \frac{\langle u_{j\mathbf{k}}|v_x|u_{j'\mathbf{k}}\rangle\langle u_{j'\mathbf{k}}|v_x|u_{j\mathbf{k}}\rangle}{(\epsilon_{j\mathbf{k}}-\epsilon_{j'\mathbf{k}})^2}. \qquad (4.28)$$

While in Sec. 4.2.3 this was integrated in the BZ, for a metal the $\mathbf{k}$ integral is limited to the volume inside the Fermi surface. Nonetheless, the transformation of the integrand, from a sum over states into a curvature, proceeds in the same way. The geometrical contribution to the AHE is therefore proportional to the integral of the Berry curvature within the Fermi surface. To allow for band crossings, we write the intrinsic AHE conductivity in dimension d as

$$\sigma_{xy} = 2\frac{e^2}{\hbar}\frac{1}{(2\pi)^d}\mathrm{Im}\sum_{j=1}^{\infty}\int_{\epsilon_{j\mathbf{k}}<\epsilon_\mathrm{F}} d\mathbf{k}\, \langle \partial_x u_{j\mathbf{k}}|\partial_x u_{j\mathbf{k}}\rangle; \qquad (4.29)$$

the analogy to Eqs. (4.20) and (4.21) is self evident. But this analogy is partly misleading: at variance with the quantum Hall case, no macroscopic $\mathbf{B}$ field is present here, and the u orbitals in Eq. (4.29) are the (periodic parts of) genuine Bloch orbitals.

Given that the Fermi surface is symmetrical under $\mathbf{k} \to -\mathbf{k}$, the symmetry considerations of Sec. 3.7 show that Eq. (4.29) can be nonzero only if time-reversal symmetry is broken, while inversion symmetry is irrelevant. The typical case studies are the ferromagnetic metals, whose ground state breaks indeed time-reversal symmetry in *absence* of a macroscopic $\mathbf{B}$ field.

First-principle calculations were performed for Ni, Cu, and Fe, as well as or for some oxides. The intrinsic geometric contribution appears to be the dominant one. These calculations also pointed out the crucial role played by avoided crossings of the bands near the Fermi surface, which induce a very spiky behavior of the Berry curvature in the BZ. More than 10^6 $\mathbf{k}$ points where used in Ref. [71] in order to perform the integration in Eq. (4.29); a more efficient strategy was devised later [72].

A noninteracting (e.g. KS) many-electron system is a trivial example of a Fermi liquid. Haldane [73] pointed out that the very basic tenet of Landau's Fermi-liquid theory is that charge transport involves only quasiparticles

with energies within $k_{\rm B}T$ of the Fermi level. This is apparently at odds with Eq. (4.29), which is an integration over the whole occupied Fermi sea. The two viewpoints can be reconciled, essentially via an integration by parts [73]. Even this alternative form has been implemented in first-principle calculations [74].

4.5 Semiclassical transport

The semiclassical theory of Bloch electron dynamics plays a fundamental role in the physics of metals and semiconductors, and is a typical textbook topic [75]. The theory addresses the motion of a wave packet built as a superposition of Bloch states from the n-th band

$$|W\rangle = \int_{\rm BZ} d\mathbf{k}\; a(\mathbf{k},t)|\psi_{n\mathbf{k}}\rangle, \tag{4.30}$$

where the envelope function is well localized in $\mathbf{k}$-space. Because of this, it is delocalized in $\mathbf{r}$ space; we assume, however, that its center of mass is well defined. Owing to this, we may define the wave vector $\mathbf{k}$ and the center $\mathbf{r}$ of the wave packet as

$$\mathbf{k} = \int_{\rm BZ} d\mathbf{k}'\; \mathbf{k}'\, |a(\mathbf{k}',t)|^2; \qquad \mathbf{r} = \langle W|\mathbf{r}|W\rangle. \tag{4.31}$$

4.5.1 Textbook equations of motion

In absence of collisions, the equations of motion reported in textbooks and routinely used in device engineering are

$$\begin{aligned} \dot{\mathbf{r}} &= \frac{1}{\hbar}\frac{\partial \epsilon_{\mathbf{k}}}{\partial \mathbf{k}} \\ \hbar\dot{\mathbf{k}} &= -e\,(\mathbf{E} + \frac{1}{c}\dot{\mathbf{r}} \times \mathbf{B}), \end{aligned} \tag{4.32}$$

where $\epsilon_{\mathbf{k}}$ is the band structure of the relevant band, and $\mathbf{E}$ and $\mathbf{B}$ are the perturbing fields, assumed weak and slowly varying in space and time.

As emphasized in Ref. [75], the derivation of Eq. (4.32), despite the formal simplicity of the result, is “a formidable task”. The early derivations date from the 1930s; the problem was reconsidered several times in the literature, by Slater [76], Luttinger [77], and Zak [78] among others.

4.5.2 Modern equations of motion

The ultimate analysis of semiclassical transport owes to a couple of papers by Q. Niu and coworkers [60, 61] (see also Ref. [12]); a couple of terms, detailed below, are missing in Eq. (4.32).

The band structure acquires a correction due to the orbital moment $\mathbf{m}(\mathbf{k})$ of the wave packet:

$$\epsilon_{\mathbf{k}} \rightarrow \tilde{\epsilon}_{\mathbf{k}} - \mathbf{m}(\mathbf{k})\cdot\mathbf{B} \qquad \mathbf{m}(\mathbf{k}) = -\frac{ie}{2\hbar c}\langle \nabla_{\mathbf{k}} u_{\mathbf{k}}| \times (H_{\mathbf{k}} - \epsilon_{\mathbf{k}})\,|\nabla_{\mathbf{k}} u_{\mathbf{k}}\rangle. \tag{4.33}$$

Furthermore, the canonical momentum $\hbar\mathbf{k}$ has to be replaced with the kinetic momentum, which includes the (geometrical) vector potential. In the Newton-like equation of motion, its contribution is reminiscent of a Lorentz force in reciprocal space. Eq. (4.32) must then be replaced with

$$\begin{aligned} \dot{\mathbf{r}} &= \frac{1}{\hbar}\frac{\partial \tilde{\epsilon}_{\mathbf{k}}}{\partial \mathbf{k}} - \dot{\mathbf{k}} \times \mathbf{\Omega}(\mathbf{k}) \\ \hbar\dot{\mathbf{k}} &= -e\,(\mathbf{E} + \frac{1}{c}\dot{\mathbf{r}} \times \mathbf{B}), \end{aligned} \tag{4.34}$$

where $\mathbf{\Omega}(\mathbf{k})$ is the Berry curvature of the relevant band, having the dimensions of a squared length. Notice that the curvature is nonzero even in presence of time-reversal symmetry, provided that the crystal is noncentrosymmetric.

4.5.3 Geometrical correction to the density of states

In a remarkable 2005 paper by Xiao, Shi, and Niu [79] it was pointed out that Eq. (4.34), in presence of a nonzero $\mathbf{B}$ field, violates Liouville's theorem. This means that the volume element $\Delta V = \Delta\mathbf{r}\Delta\mathbf{k}$ changes in time during the evolution of the system; it is possible, however to remedy this shortcoming in an elegant way.

The standard volume element of the phase space in dimension d is $\Delta\mathbf{r}\Delta\mathbf{p}/h^d = (2\pi)^{-d}\Delta\mathbf{r}\Delta\mathbf{k}$. According to Ref. [79] this has to be modified by a geometrical term as:

$$\frac{1}{(2\pi)^d}\Delta\mathbf{r}\Delta\mathbf{k} \rightarrow \frac{1}{(2\pi)^d}\left(1 + \frac{2\pi}{\phi_0}\mathbf{B}\cdot\mathbf{\Omega}(\mathbf{k})\right)\Delta\mathbf{r}\Delta\mathbf{k}, \tag{4.35}$$

where ϕ_0 is the flux quantum.

As a consequence of the modified density of states, the Fermi volume of a metal changes when a macroscopic $\mathbf{B}$ field is switched on at constant electron density. If instead we keep the chemical potential μ constant, then the electron density n depends on $\mathbf{B}$. At zero temperature

$$\begin{aligned} n &= \frac{1}{(2\pi)^d}\int_{\mathrm{BZ}} d\mathbf{k}\,\left(1 + \frac{2\pi}{\phi_0}\mathbf{B}\cdot\mathbf{\Omega}(\mathbf{k})\right)\vartheta(\mu - \epsilon_{\mathbf{k}}) \\ \left(\frac{\partial n}{\partial \mathbf{B}}\right)_\mu &= \frac{1}{(2\pi)^{d-1}\phi_0}\int_{\mathrm{BZ}} d\mathbf{k}\;\mathbf{\Omega}(\mathbf{k})\,\vartheta(\mu - \epsilon_{\mathbf{k}}) \end{aligned} \tag{4.36}$$

The latter assumes a perspicuous meaning in 2d, when μ is in a gap:

$$\left(\frac{\partial n}{\partial B}\right)_\mu = \frac{1}{2\pi\phi_0}\int_{\mathrm{BZ}} d\mathbf{k}\;\mathbf{\Omega}(\mathbf{k}) = \frac{C_1}{\phi_0} = -\frac{1}{ec}\sigma_{xy}. \tag{4.37}$$

For a quantum Hall system, this goes under the name of Streda formula [80], and had been first derived in 1982 in a very different way.

4.6 Quantum transport

4.6.1 Transport by a single state

We are going to study here the current induced by an adiabatic change of the potential, or more generally of the Hamiltonian, in the single-particle case. We

indicate as $\psi_n(t)$ the adiabatic instantaneous eigenstates, and with $\psi(t)$ the time evolution of the ground state. In order to get rid of the dynamical phase, it is better to deal with density matrices

$$\rho(t) = |\psi(t)\rangle\langle\psi(t)| = |\psi_0(t)\rangle\langle\psi_0(t)| + \Delta\rho(t). \tag{4.38}$$

The velocity of this state is

$$\begin{aligned} \mathbf{v}(t) &= \mathrm{Tr}\ \{\rho(t)\,\mathbf{v}\} \qquad\qquad (4.39)\\ &= \langle\psi_0(t)|\,\mathbf{v}\,|\psi_0(t)\rangle + \sum_n \langle\psi_0(t)|\,\Delta\rho(t)\,|\psi_n(t)\rangle\langle\psi_n(t)|\,\mathbf{v}\,|\psi_0(t)\rangle. \end{aligned}$$

Since the adiabatic density matrix commutes with $H(t)$, the time evolution is

$$\begin{aligned} [H(t), \Delta\rho(t)] &= i\hbar\dot{\rho}(t) \simeq i\hbar\frac{d}{dt}|\psi_0(t)\rangle\langle\psi_0(t)| \\ &= i\hbar\,(\,|\dot{\psi}_0(t)\rangle\langle\psi_0(t)| - |\psi_0(t)\rangle\langle\dot{\psi}_0(t)|\,), \end{aligned} \tag{4.40}$$

where we are neglecting a term of higher order in the adiabaticity parameter. We now take the matrix elements between $\langle\psi_0(t)|$ and $|\psi_n(t)\rangle$:

$$(E_0 - E_n)\langle\psi_0(t)|\,\Delta\rho(t)\,|\psi_n(t)\rangle = i\hbar\,(\,\langle\psi_n(t)|\dot{\psi}_0(t)\rangle - \langle\dot{\psi}_0(t)|\psi_n(t)\rangle\,). \tag{4.41}$$

The term with $n = 0$ in the rhs vanishes because of norm conservation; replacement into Eq. (4.39) yields

$$\begin{aligned} \mathbf{v}(t) &= \langle\psi_0(t)|\,\mathbf{v}\,|\psi_0(t)\rangle \\ &\quad + i\hbar\sum_{n\neq 0}\left[\frac{\langle\dot{\psi}_0(t)|\psi_n(t)\rangle\langle\psi_n(t)|\,\mathbf{v}\,|\psi_0(t)\rangle}{E_0 - E_n} - \text{c.c.}\right]. \end{aligned} \tag{4.42}$$

The first term on the rhs is zero in the special case where—at all times—the Hamiltonian $H(t)$ is time-reversal invariant and the state is nondegenerate.

4.6.2 Current carried by filled bands

We now exploit the previous result for a system of noninteracting electrons in the case where the Hamiltonian $H(t)$ is lattice periodical and the ground state is insulating; this means that the gap remains finite at all t. It will be enough to consider the simple case of just one filled band, with band index zero; the current carried by each Bloch orbital $|\psi_{0\mathbf{k}}\rangle = \mathrm{e}^{i\mathbf{k}\cdot\mathbf{r}}|u_{0\mathbf{k}}\rangle$ is

$$\begin{aligned} \mathbf{v}_{\mathbf{k}}(t) &= \langle\psi_{0\mathbf{k}}|\,\mathbf{v}\,|\psi_{0\mathbf{k}}\rangle + i\hbar\sum_{j\neq 0}\left[\frac{\langle\dot{\psi}_{0\mathbf{k}}|\psi_{j\mathbf{k}}\rangle\langle\psi_{j\mathbf{k}}|\,\mathbf{v}\,|\psi_{0\mathbf{k}}\rangle}{\epsilon_{0\mathbf{k}} - \epsilon_{j\mathbf{k}}} - \text{c.c.}\right] \\ &= \langle u_{0\mathbf{k}}|\,\mathbf{v}\,|u_{0\mathbf{k}}\rangle + i\hbar\sum_{j\neq 0}\left[\frac{\langle\dot{u}_{0\mathbf{k}}|u_{j\mathbf{k}}\rangle\langle u_{j\mathbf{k}}|\,\mathbf{v}\,|u_{0\mathbf{k}}\rangle}{\epsilon_{0\mathbf{k}} - \epsilon_{j\mathbf{k}}} - \text{c.c.}\right], \end{aligned} \tag{4.43}$$

where the t dependence of the rhs is now implicit. We then adopt the usual formula for the velocity, Eqs. (1.18) and (4.18), and the analog of Eq. (3.27):

$$\mathbf{v} = \frac{1}{\hbar}\nabla_{\mathbf{k}}H_{\mathbf{k}}, \quad |\nabla_{\mathbf{k}}u_{0\mathbf{k}}\rangle = \frac{1}{\hbar}\sum_{j\neq 0}|u_{j\mathbf{k}}\rangle\frac{\langle u_{j\mathbf{k}}|\,\mathbf{v}\,|u_{0\mathbf{k}}\rangle}{\epsilon_{0\mathbf{k}} - \epsilon_{j\mathbf{k}}} \tag{4.44}$$

$$v_{\mathbf{k}}(t) = \frac{1}{\hbar}\nabla_{\mathbf{k}}\epsilon_{0\mathbf{k}} + i\,(\,\langle \dot{u}_{0\mathbf{k}}|\nabla_{\mathbf{k}} u_{0\mathbf{k}}\rangle - \langle \nabla_{\mathbf{k}} u_{0\mathbf{k}}|\dot{u}_{0\mathbf{k}}\rangle\,). \tag{4.45}$$

The first term on the rhs integrates to zero over the BZ, while the second is clearly a Berry curvature component in the four-dimensional $\mathbf{k}, t$ domain. The current density carried by a filled band in dimension d is

$$\mathbf{j}(t) = -\frac{ie}{(2\pi)^d}\int_{\mathrm{BZ}} d\mathbf{k}\;(\,\langle \dot{u}_{0\mathbf{k}}|\nabla_{\mathbf{k}} u_{0\mathbf{k}}\rangle - \langle \nabla_{\mathbf{k}} u_{0\mathbf{k}}|\dot{u}_{0\mathbf{k}}\rangle\,); \tag{4.46}$$

the Bloch states are normalized to one in the crystal cell (as everywhere in the present Notes).

4.6.3 Quantization of charge transport

Let us consider the special case of a simple cubic crystal with lattice constant a. The transported charge in time T in the z direction across one cell is

$$\begin{aligned} Q &= \int_0^T dt\; I_z(t) = a^2 \int_0^T dt\; j_z(t) = -\frac{ea^2}{(2\pi)^2}\int_{-\pi/a}^{\pi/a} dk_x \int_{-\pi/a}^{\pi/a} dk_y \\ &\times \frac{i}{2\pi}\int_0^T dt \int_{-\pi/a}^{\pi/a} dk_z\;(\,\langle \dot{u}_{0\mathbf{k}}|\partial_z u_{0\mathbf{k}}\rangle - \langle \partial_z u_{0\mathbf{k}}|\dot{u}_{0\mathbf{k}}\rangle\,). \end{aligned} \tag{4.47}$$

If the time evolution of the Hamiltonian is cyclic $H(T) = H(0)$, then the second line in Eq. (4.47) is clearly a Chern number C (in the k_z, t variables) and is integer. Notice also that C is dimensionless, and therefore does not depend on how fast the Hamiltonian varies with time; ideally, the adiabatic regime means $T \to \infty$.

We arrive therefore at the outstanding result

$$Q = \int_0^T dt\; I_z(t) = e \times \text{ integer} \tag{4.48}$$

first proved by Thouless in 1983 [81]; it holds of course for any dimension d. Let me restate the theorem: if the Hamiltonian is changed adiabatically in such a way that it returns to its starting value in time T, the transported charge in an infinite periodic system is quantized provided that the system remains insulating at all times. A cycle pumps an integer number of elementary charges across the system.

Among the examples which realize a "Thouless pump", the original paper suggests a sliding charge-density wave. A more outstanding manifestation of quantized charge transport was pointed out shortly afterwards by Pendry and Hodges [82]: Faradays' laws of of electrolysis (1832). The mass/charge transfer ratio shows that charge is always transported in units of e per ion, to the extent that electrolytic cells are used as standards of current. If a given ion sits at one electrode at $t = 0$, and if it drifts to the other one at $t = T$, the Hamiltonian can be considered as cyclic, whence charge quantization follows. However, at intermediate t values the charge "belonging" to a given ion is definitely non quantized, and arbitrarily defined: for a review of the possible definitions, see Ref. [83].

Thouless quantization of charge transport [81], discussed above, also has profound relationships to later advances: namely, to the topological explanation of the quantization of surface charge (discussed in Sec. 5.3.4), and to the modern theory of polarization (discussed in Sec. 5.3).

Chapter 5

Modern theory of polarization

The macroscopic polarization **P** is a fundamental concept that all undergraduates learn about in elementary courses [84, 85]. In view of this, it is truly extraordinary that until rather recently there was no generally accepted formula for **P** in condensed matter, even as a matter of principles. **P** is an intensive vector quantities that intuitively carries the meaning of dipole per unit volume. Most textbooks [75, 86] provide a flawed definition of **P**, not implementable in practical computations [87].

A genuine change of paradigm was initiated by a couple of important papers [88, 89], after which the major development was introduced by King-Smith and Vanderbilt in 1992 (paper published in 1993 [90]). Other important advances occurred during the 1990s [91, 92] and the so-called "modern theory of polarization"it is at a mature stage since more than a decade; several reviews have appeared in the literature over the years [1, 2, 3, 5, 6, 7, 8].

Aiming at a computational physics readership, it is worth emphasizing that most ab-initio electronic-structure codes on the market, for dealing with either crystalline or noncrystalline materials, implement the modern theory of polarization as a standard option. A nonexaustive list includes ABINIT [56], CRYSTAL [57], QUANTUM-ESPRESSO [59], SIESTA [93], VASP [58], and CPMD [94]. Implementations of the modern theory have been instrumental since more than a decade in the study of ferroelectric and piezoelectric materials [95, 96, 97].

The basic concepts of the modern theory of polarization also start reaching a few textbooks [98], though very slowly; most of them are still plagued with erroneous concepts and statements.

5.1 Polarization and fields

The modern theory of polarization, at least in its original form, only addresses the polarization **P** in a null macroscopic **E** field; in this case **P** can be nonzero only if the medium breaks inversion symmetry. This applies to noncentrosymmetric crystals, and more generally to cases where inversion symmetry is broken by some perturbation (as e.g. a frozen phonon, or piezoelectric strain).

It must be realized that, insofar as we address an infinite system with no boundaries, the **E** field is quite arbitrary. The microscopic charge density is neutral in average and lattice periodical; the value of **E** is just an arbitrary boundary condition for the integration of Poisson's equation. The usual choice

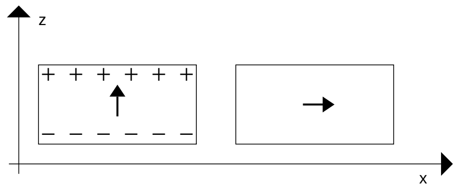

Figure 5.1: Macroscopic polarization $\mathbf{P}$ in a slab normal to z, for a vanishing external field $\mathbf{E}^{(\mathrm{ext})}$. Left: When $\mathbf{P}$ is normal to the slab, a depolarizing field $\mathbf{E} = -4\pi\mathbf{P}$ is present inside the slab, and charges at its surface, with areal density $\sigma_{\mathrm{surface}} = \mathbf{P} \cdot \mathbf{n}$ Right: When $\mathbf{P}$ is parallel to the slab, no depolarizing field and no surface charge is present.

(performed within all electronic-structure codes) is to impose a lattice-periodical Coulomb potential, i.e. $\mathbf{E} = 0$. Imposing a given nonzero value of $\mathbf{E}$ is equally legitimate (in insulators), although technically more difficult [99, 100]).

When addressing a finite sample with boundaries, the $\mathbf{E}$ field is in principle measurable inside the material, without reference to what happens at the sample boundary; this is not the case of $\mathbf{D}$. In fact, $\mathbf{E}$ obtains by averaging over a macroscopic length scale the microscopic electric field $\mathbf{E}^{(\mathrm{micro})}(\mathbf{r})$, which fluctuates at the atomic scale [85]. In a macroscopically homogeneous system the macroscopic field $\mathbf{E}$ is constant, and in crystalline materials it coincides with the cell average of $\mathbf{E}^{(\mathrm{micro})}(\mathbf{r})$. A lattice-periodical potential enforces $\mathbf{E} = 0$; for a supercell calculation, this applies to the field average over the supercell, while in different regions there can be a nonzero macroscopic field.

As explained so far, there is no need of addressing finite samples and external vs. internal fields from a theoretician's viewpoint. Nonetheless a brief digression is in order, given that experiments *are* performed over finite samples, often in external fields. Suppose a finite macroscopic sample is inserted in a constant external field $\mathbf{E}^{(\mathrm{ext})}$: the microscopic field $\mathbf{E}^{(\mathrm{micro})}(\mathbf{r})$ coincides with $\mathbf{E}^{(\mathrm{ext})}$ far away from the sample, while it is different inside because of screening effects. If we choose an homogeneous sample of *ellipsoidal shape*, then the macroscopic average of $\mathbf{E}^{(\mathrm{micro})}(\mathbf{r})$, i.e. the macroscopic screened field $\mathbf{E}$, is constant in the bulk of the sample. The shape effects are embedded in the depolarization coefficients [84]: the simplest case is the extremely oblate ellipsoid, i.e. a slab of a macroscopically homogeneous dielectric; more details are given in Ref. [8]. For the slab geometry in a vanishing external field $\mathbf{E}^{(\mathrm{ext})}$ the internal field $\mathbf{E}$ vanishes when $\mathbf{P}$ is parallel to the slab (transverse polarization), while $\mathbf{E} = -4\pi\mathbf{P}$ is the depolarization field when $\mathbf{P}$ is normal to the slab (longitudinal polarization): see Fig. 5.1.

5.2 Polarization "itself" vs. polarization difference

Novel ideas about macroscopic polarization emerged in the early 1990s [88, 89]; these led to the modern theory, based on a Berry phase, which was founded by King-Smith and Vanderbilt soon afterwards [90]. At its foundation, the modern theory was limited to a crystalline system in an independent-electron

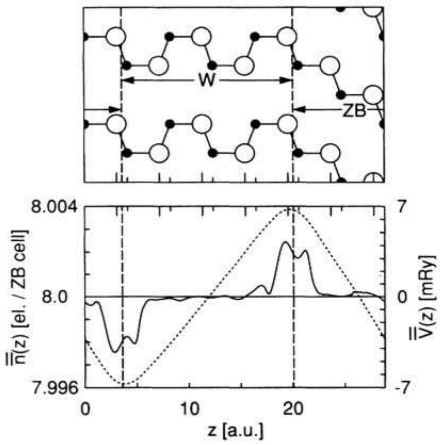

Figure 5.2: Top panel: The 14-atom BeO supercell in a vertical plane through the BeO bonds; the wurtzite (W) and zincblende (ZB) stackings are perspicuous. Bottom panel: Macroscopic averages of the valence electron density (solid) and of the electrostatic potential (dotted).

framework (either KS or Hartree-Fock). Later, the theory was extended to correlated and/or disordered systems [91, 92].

The first calculation ever of spontaneous polarization was published in 1990 [88]. The case study was BeO: it has the simplest structure where inversion symmetry is absent (i.e. wurtzite), and furthermore its constituents are first-row atoms. The idea was to address the macroscopic polarization of a slab of finite thickness, with faces normal to the c axis, embedding it in an *ad hoc* medium which (i) has no bulk polarization for symmetry reasons, and (2) does not produce any geometrical or chemical perturbation at the interface. The optimal choice is a fictitious BeO in the zincblende structure. Because of obvious reasons, the system is periodically replicated in a supercell geometry (Fig. 5.2, top panel). The selfconsistent calculation shows well localized interface charges, of opposite sign and equal magnitudes at the two nonequivalent interfaces (Fig. 5.2, bottom panel). The interface charge is related to the difference in polarization between the two materials: $\sigma_{\text{interface}} = \Delta\mathbf{P} \cdot \mathbf{n}$. The computer experiment provides the value of $\sigma_{\text{interface}}$, and since $\mathbf{P}$ vanishes by symmetry in the zincblende slab, one thus obtains the bulk value of $\mathbf{P}$ in the wurtzite material. Notice that here $\mathbf{P}$ is a *longitudinal* polarization, in a depolarizing field.

It must be emphasized that the quantity really "measured" in this computer experiment is $\Delta\mathbf{P}$, *not* the polarization $\mathbf{P}$ itself. After Ref [88] was published, a study of the experimental literature showed that—contrary to an incorrect widespread belief—no experimental value of $\mathbf{P}$ in any wurtzite material exists: only estimates are available. Ref. [88] marks, as said above, a change of paradigm: polarization must be *defined* by means of *differences*, and the concept of polarization "itself" must be abandoned. With hindsight, it is nowadays pretty clear that the problem, as well as its solution, exists already at the classical level: this is sketched in Fig. 5.3. Most textbooks are missing this very basic fact.

The modern theory avoids addressing the "absolute" polarization of a given equilibrium state, quite in agreement with the experiments, which invariably measure polarization *differences*. Instead, the theory addresses differences in polarization between two states of the material that can be connected by an adiabatic switching process. The time-dependent Hamiltonian is assumed to

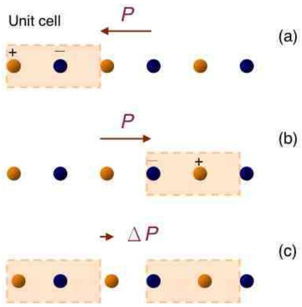

Figure 5.3: A 1d solid with infinite length. Different choices of the unit cell give different $\mathbf{P}$ values: (a), (b). On the other hand, the change of polarization $\Delta\mathbf{P}$ does not depend on the choice of the unit cell (c).

remain insulating at all times, and the polarization difference is then *defined* [89] as the time-integrated transient macroscopic current that flows through the insulating sample during the switching process:

$$\Delta\mathbf{P} = \mathbf{P}(\Delta t) - \mathbf{P}(0) = \int_0^{\Delta t} dt\, \mathbf{j}(t). \tag{5.1}$$

In the adiabatic limit $\Delta t \to \infty$ and $\mathbf{j}(t) \to 0$, while $\Delta\mathbf{P}$ stays finite. Addressing currents (instead of charges) explains the occurrence of *phases* of the wavefunctions (instead of square moduli) in the modern theory. Eventually the time integration in Eq. (5.1) will be eliminated, leading to a two-point formula involving only the initial and final states.

5.3 Independent electrons

5.3.1 The King-Smith and Vanderbilt formula

For a crystalline system of independent electrons the expression for the transient current occurring in Eq. (5.1) is precisely the same as previously derived for quantum transport, Eq. (4.46). Therefore for one band (index 0) and single occupancy in dimension d the electronic contribution to the polarization difference is

$$\Delta\mathbf{P} = -\frac{ie}{(2\pi)^d} \int_0^{\Delta t} dt \int_{\mathrm{BZ}} d\mathbf{k}\, (\, \langle \dot{u}_{0\mathbf{k}} | \nabla_{\mathbf{k}} u_{0\mathbf{k}} \rangle - \langle \nabla_{\mathbf{k}} u_{0\mathbf{k}} | \dot{u}_{0\mathbf{k}} \rangle\,); \tag{5.2}$$

the (classical) nuclear contribution must be added separately. We remind that crystal-cell neutrality is essential. Notice also that, given the occurrence of Bloch states, the Hamiltonian is lattice periodical at all t: this implicitly means that in Eq. (5.2) $\Delta\mathbf{P}$ is evaluated at $\mathbf{E} = 0$.

It is now expedient to introduce a dimensionless adiabatic time λ, with $|u_{j\mathbf{k}}(t)\rangle = |u_{j\mathbf{k}}(\lambda(t))\rangle$, $\lambda(0) = 0$, $\lambda(1) = \Delta t$. Eq. (5.2) becomes then, for n doubly occupied bands (index $j = 1, n$) in 3d:

$$\begin{aligned} \Delta\mathbf{P} &= \mathbf{P}(1) - \mathbf{P}(0) = \int_0^1 d\lambda\, \partial_\lambda \mathbf{P}(\lambda) \\ \partial_\lambda \mathbf{P} &= -\frac{2ie}{(2\pi)^3} \sum_{j=1}^{n} \int_{\mathrm{BZ}} d\mathbf{k}\, (\, \langle \partial_\lambda u_{j\mathbf{k}} | \nabla_{\mathbf{k}} u_{j\mathbf{k}} \rangle - \langle \nabla_{\mathbf{k}} u_{j\mathbf{k}} | \partial_\lambda u_{j\mathbf{k}} \rangle\,). \end{aligned} \tag{5.3}$$

It is essential that the gap does not close, i.e. the system remain insulating, for all λ values.

The expression in Eq. (5.3) can be integrated with respect to λ to obtain

$$\mathbf{P}(\lambda) = -\frac{2ie}{(2\pi)^3} \sum_{j=1}^{n} \int_{\mathrm{BZ}} d\mathbf{k}\, \langle u_{j\mathbf{k}} | \nabla_{\mathbf{k}} u_{j\mathbf{k}} \rangle : \tag{5.4}$$

this is the (by now famous) King-Smith and Vanderbilt formula [90], yielding the polarization of the final state minus the polarization of the initial state, Eq. (5.3). To understand the meaning of the $\mathbf{k}$ integral in 3d we take the simple example of a simple cubic lattice of constant a, similarly to Eq. (4.47):

$$P_z(\lambda) = -\frac{2e}{(2\pi)^3} \sum_{j=1}^{n} \int_{-\pi/a}^{\pi/a} dk_x \int_{-\pi/a}^{\pi/a} dk_y \left[i \int_{-\pi/a}^{\pi/a} dk_z \langle u_{j\mathbf{k}} | \partial_{k_z} u_{j\mathbf{k}} \rangle \right], \tag{5.5}$$

where the square parenthesis highlights the Berry-phase, to be compared to Eq. (3.43).

In first-principle implementations, the Berry phase is discretized as in Eqs. (3.45) and (3.46), and the remaining 2d $\mathbf{k}$-integral is discretized in the trivial way. The first calculation ever of the "spontaneous" polarization of a ferroelectric material ($KNbO_3$) appeared in 1993 [101], and agreed within 10% with the measured values. As said above, the Berry-phase formula is nowadays implemented in most first-principle codes.

5.3.2 The quantum of polarization

Given that every phase is defined modulo 2π, all of the two-point formulas for $\Delta\mathbf{P}$ in terms of Berry phases are arbitrary modulo a polarization "quantum". This is the tradeoff one has to pay when switching from the curvature formula, Eq. (5.3)—where no such arbitrariness exists—to the two-point King-Smith and Vanderbilt formula, where only the connection occurs. The actual arbitrariness of $\Delta\mathbf{P}$ in 3d is $2e\mathbf{R}/V_{\mathrm{cell}}$, where $\mathbf{R}$ is a lattice vector and V_{cell} is the cell volume (the 2 factor owes to double band occupancy). A similar arbitrariness of an integer times $e\mathbf{R}/V_{\mathrm{cell}}$ occurs for the classical nuclear contribution to polarization.

The quantum arbitrariness is rarely a problem in practice. In most cases, the change in $\mathbf{P}$ that can be induced by a perturbation, such as a small sublattice displacement, is insufficient to cause $\mathbf{P}$ to change by a large fraction of the quantum. Where exceptions exist the ambiguity can be resolved by subdividing the adiabatic path into several shorter intervals, for each of which the change in $\mathbf{P}$ is unambiguous for practical purposes. Additional problems may occur in the discretized version of the Berry-phase formula; this is discussed e.g. in Ref. [8] and, in more detail, in Ref. [7].

Here we stress that the quantum ambiguity is an essential aspect of the theory. For example, for the case of a *closed cyclic* adiabatic evolution of the system, in which the parameter values $\lambda = 0$ and $\lambda = 1$ label the *same physical state* of the system, we retrieve the quantization of charge transport, discussed in Sec. 4.6, and governed by a Chern number.

5.3.3 Wannier functions

The KS (or Hartee-Fock) ground state is a Slater determinant of doubly occupied orbitals; any unitary transformation of the occupied states among themselves leaves the determinantal wavefunction invariant (apart for an irrelevant phase factor), and hence it leaves invariant any KS ground-state property.

For an insulating crystal the Bloch KS orbitals of completely occupied bands can be transformed to localized Wannier orbitals (or functions) WFs. This is known since 1937 [102], but for many years the WFs have been mostly used as a formal tool; they became a popular topic in computational electronic structure only after the seminal work of Marzari and Vanderbilt [103]. A comprehensive review appeared as Ref. [18], and a public-domain implementation is in WANNIER90 [104]. If the crystal is metallic, the WFs can still be technically useful [105], but it must be emphasized that the ground state *cannot* be written as a Slater determinant of localized orbitals of any kind, as a matter of principle [106].

The transformation of the Berry phase formula in terms of WFs provides an alternative, and perhaps more intuitive, viewpoint. The formal transformation was known since the 1950s [107], although the physical meaning of the formalism was not understood until the advent of the modern theory of polarization.

The unitary transformation which defines the WF $w_{j\mathbf{R}}(\mathbf{r})$, labeled by band j and unit cell $\mathbf{R}$, within our normalization is

$$|w_{j\mathbf{R}}\rangle = \frac{V_{\rm cell}}{(2\pi)^3} \int_{\rm BZ} d\mathbf{k}\; \mathrm{e}^{i\mathbf{k}\cdot\mathbf{R}}\; |\psi_{j\mathbf{k}}\rangle \; . \tag{5.6}$$

If one then defines the "Wannier centers" as $\mathbf{r}_{j\mathbf{R}} = \langle w_{j\mathbf{R}}|\mathbf{r}|w_{j\mathbf{R}}\rangle$, it is rather straightforward to prove that Eq. (5.4) is equivalent to

$$\mathbf{P}^{(\rm el)}(\lambda) = -\frac{2e}{V_{\rm cell}} \sum_{j=1}^{n} \mathbf{r}_{j\mathbf{0}}. \tag{5.7}$$

This means that the electronic term in the macroscopic polarization $\mathbf{P}$ is (twice) the dipole of the Wannier charge distributions in the central cell, divided by the cell volume. The nuclear term is obviously similar in form to Eq. (5.7); the sum of both terms is charge neutral.

WFs are severely gauge-dependent, since the phases of the $|\psi_{j\mathbf{k}}\rangle$ appearing in Eq. (5.6) can be chosen arbitrarily. However, their centers are gauge-invariant modulo a lattice vector. Therefore $\mathbf{P}^{(\rm el)}$ in Eq. (5.7) is affected by the same "quantum" indeterminacy discussed above. We also stress, once more, that Eq. (5.7)—as well as Eq. (5.3)—is to be used in polarization *differences*, and does not define polarization itself.

The modern theory, when formulated in terms of WFs, becomes much more intuitive, and in a sense vindicates the venerable Clausius-Mossotti viewpoint [108]: in fact, the charge distribution is partitioned into localized contributions, each providing an electric dipole, and these dipoles yield the electronic term in $\mathbf{P}$. However, it is clear from Eq. (5.6) that the *phase* of the Bloch orbitals is essential to arrive at the right partitioning. Any decomposition based on charge only is severely nonunique and does not provide in general the right $\mathbf{P}$, with the notable exception of the extreme case of molecular crystals.

In the latter case, in fact, we may consider the set of WFs centered on a given molecule; their total charge distribution coincides—in the weakly interacting limit—with the electron density of the isolated molecule (possibly in a local field). This justifies the elementary Clausius-Mossotti viewpoint. It is worth mentioning that the dipole of a polar molecule is routinely computed in a supercell geometry via the single-point Berry phase discussed below [109]. The dipole value coincides with the one computed in the trivial way in the large supercell limit. Finite-size corrections, due to the local field (different in the two cases), can also be applied [110].

The case of alkali halides—where the model is often phenomenologically used—deserves a different comment [8]. The electron densities of isolated ions (with or without fields) are quite different from the corresponding WFs charge distributions, for instance because of orthogonality constraints: hence the Clausius-Mossotti model is *not* justified in its elementary form, despite contrary statements in the literature. For a detailed analysis, see Ref. [111].

5.3.4 The surface charge theorem

The early occurrences of the theorem of quantization of the surface charge [32, 33, 34, 35] are discussed above, Sec. 2.3. The topological nature of this theorem was first realized by Niu [112] in 1986; here we follow the treatment of Vanderbilt and King-Smith [113] (see also Ref. [5]).

According to elementary electrostatics the macroscopic bound surface charge density σ_{surface} residing on the surface of a sample is related to the polarization in the interior by $\sigma_{\text{surface}} = \hat{\mathbf{n}} \cdot \mathbf{P}$, where $\hat{\mathbf{n}}$ is the surface normal. One defines the bound charge σ_{surface} by saying that no free charge is present, but what, precisely, does this mean? The surface must be insulating, with the electron chemical potential lying in a gap that is common to both bulk and surface. But this is not a unique prescription, since there can be a surface band which is entirely occupied or entirely empty. The two cases differ by a polarization quantum in the corresponding $\mathbf{P}$ value. In fact, given that the bulk polarization $\mathbf{P}$ is arbitrary modulo $e\mathbf{R}/V_{\text{cell}}$, it follows that the charge per surface area is defined modulo a quantum

$$\sigma_{\text{surface}} = \hat{\mathbf{n}} \cdot \mathbf{P} \quad \text{modulo } \frac{e}{A_{\text{surface}}} \, . \tag{5.8}$$

An equivalent formulation of the surface charge theorem can be arrived at by means of WFs. The WF approach is most perspicuous for quasi-1d systems (e.g. insulating polymers); a pedagogical presentation is in Ref. [114]. Notice that there cannot be surface states in a polymer: all terminations are "insulating".

The bulk-surface correspondence encoded in Eq. (5.8) is an outstanding manifestation of topology in condensed matter physics: the surface charge of an insulating surface is "topologically protected". The actual value of σ_{surface} among the discrete allowed values is then determined by energy considerations.

For a centrosymmetric crystal it is tempting to guess that $\mathbf{P}(\lambda)$ in the Berry-phase formula, Eq. (5.4), vanishes for any λ. Instead, this is not the case: centrosymmetry only dictates that $\mathbf{P} = -\mathbf{P}$ modulo a quantum, and therefore Eq. (5.4) yields $\mathbf{P}$ equal to an integer or half integer multiple of the quantum $e\mathbf{R}/V_{\text{cell}}$ (for single band occupancy). Then from Eq. (5.8) it follows that the

charge per surface cell may only be an integer or half integer, as first discovered many years ago, and previously discussed in Sec. 2.3. Therein, it was observed that this important theorem is often ignored even by specialists in surface physics. A thorough analysis of polar surfaces, in the light of the present theorem, has recently been published by Stengel [115].

5.3.5 Noncrystalline systems: The single-point Berry phase

The key ingredient for computing the infrared spectrum of amorphous or liquid systems is the power spectrum of the autocorrelation function of the macroscopic polarization $\langle \mathbf{P}(t) \cdot \mathbf{P}(0) \rangle$. Since Car-Parrinello simulations are customarily performed using only $\mathbf{k} = 0$ in the (supercell) BZ, we need to analyze the single-point version of the Berry-phase formula for polarization, Eq. (5.5).

We consider a simple cubic supercell of side L; in the large L limit the BZ integral of any function $f(\mathbf{k})$ is approximated as

$$\int_{\mathrm{BZ}} d\mathbf{k}\; f(\mathbf{k}) \;\rightarrow\; \frac{(2\pi)^3}{L^3} f(0). \tag{5.9}$$

We start with the Berry phase, i.e. with the 1d integral in square parenthesis in Eq. (5.5):

$$\gamma_z = i \int_{-\frac{\pi}{L}}^{\frac{\pi}{L}} dk_z \langle u_{j\mathbf{k}} | \partial_{k_z} u_{j\mathbf{k}} \rangle = i \int_{0}^{\frac{2\pi}{L}} dk_z \langle u_{j\mathbf{k}} | \partial_{k_z} u_{j\mathbf{k}} \rangle \rightarrow -\mathrm{Im} \log \det S(\mathbf{k}_1, \mathbf{k}_2), \tag{5.10}$$

where $\mathbf{k}_1 = (0, 0, 0)$ and $\mathbf{k}_2 = (0, 0, 2\pi/L)$. In Eq. (5.10) we have used the discretized Berry phase, Eq. (3.46), with only one factor in the matrix product. Then, as in Sec. 3.9, we notice that $|u_{j\mathbf{k}_2}\rangle = \mathrm{e}^{-i\frac{2\pi z}{L}} |u_{j\mathbf{k}_1}\rangle$: therefore the overlap matrix in Eq. (5.10) becomes

$$S_{jj'}(\mathbf{k}_1, \mathbf{k}_2) = \langle u_j | \mathrm{e}^{-i\frac{2\pi z}{L}} | u_{j'} \rangle, \tag{5.11}$$

where all the orbitals $|u_j\rangle = |\psi_j\rangle$ are evaluated at $\mathbf{k} = 0$. We then approximate even the remaining integrals in Eq. (5.5) with a single point. At any time during the simulation the electronic term in the polarization is thus

$$P_z^{(\mathrm{el})}(t) = -\frac{e}{\pi L^2} \gamma_z = \frac{e}{\pi L^2} \mathrm{Im} \log \det S. \tag{5.12}$$

The nuclear (or core) contribution has a very simple form. If z_m is the instantaneous z coordinate of the m-th nucleus, and eZ_m the corresponding charge, the total polarization is

$$P_z(t) = \frac{e}{\pi L^2} \mathrm{Im} \log \det S + \frac{e}{L^3} \sum_m Z_m z_m. \tag{5.13}$$

This the expression currently used in computing power spectra and infrared spectra [116], and, more generally, whenever a single $\mathbf{k}$ point is used in the first-principle simulations [117, 109].

The polarization quantum in Eq. (5.13) is e/L^2, which vanishes in the $L \to \infty$ limit. This does not make any problem, and in fact Eq. (5.12) is routinely used

for evaluating polarization differences in noncrystalline materials. The key point is that the $L \to \infty$ limit is not actually needed; for an accurate description of a given material, it is enough to assume a *finite* L, actually larger than the relevant correlation lengths in the material. For any given length, the quantum e/L^2 sets an upper limit to the magnitude of a polarization difference accessible via the Berry phase. The larger are the correlation lengths, the smaller is the accessible $\Delta\mathbf{P}$. This is no problem at all in practice, either when evaluating static derivatives by numerical differentiation, such as e.g. in Ref. [117, 109], or when performing Car-Parrinello simulations [116]. In the latter case Δt is a Car-Parrinello time step (a few a.u.), during which the polarization varies by a tiny amount, much smaller than the quantum (the typical size of a large simulation cell nowadays is $L \simeq 50$ bohr). Whenever needed, the drawback may be overcome by splitting Δt into several smaller intervals.

5.4 Correlated wavefunctions

The treatment given so far assumes an independent-particle scheme, where polarization is evaluated as a Berry phase of one-electron orbitals, typically the KS ones. Shortly after the appearance of the King-Smith and Vanderbilt paper [90], Ortiz and Martin [91] provided the many-body generalization of the theory, where polarization is expressed as a Berry phase of the many-body wavefunction.

A subsequent development [92] provides a unified treatment of macroscopic polarization, dealing on the same footing with either independent-electron or correlated systems, and with either crystalline or disordered systems.

5.4.1 Single-point Berry phase again

Suppose we have N electrons in a cubic box of volume L^3, with a many-body Hamiltonian

$$\hat{H} = \frac{1}{2m_e} \sum_{i=1}^{N} |\mathbf{p}_i|^2 + \hat{V}, \tag{5.14}$$

and eigenfunctions $|\Psi_n\rangle$, normalized in the hypercube of volume L^{3N}. In Eq. (5.14) the potential $\hat{V}$ includes one-body and two-body (electron-electron) contributions. As usual in condensed-matter theory, we adopt periodic Born-von-Kàrmàn boundary conditions over each electron coordinate $\mathbf{r}_i$ independently, whose Cartesian components $r_{i,\alpha}$ are then equivalent to the angles $2\pi r_{i,\alpha}/L$. The potential $\hat{V}$ enjoys the same periodicity, which implies that the electric field averages to zero over the sample.

The formula for correlated wavefunctions is quite similar to Eq. (5.12), which in fact is a special case of the latter. The starting ingredient is the single-point Berry phase

$$\gamma_z = \text{Im log } \langle\Psi_0|\mathrm{e}^{i\frac{2\pi}{L}\sum_i z_i}|\Psi_0\rangle = -\text{Im log } \langle\Psi_0|\mathrm{e}^{-i\frac{2\pi}{L}\sum_i z_i}|\Psi_0\rangle, \tag{5.15}$$

and the polarization formula obtains by replacing γ_z in Eq. (5.13), or equivalently

$$(\det S)^2 \;\to\; \text{Im log } \langle\Psi_0|\mathrm{e}^{-i\frac{2\pi}{L}\sum_i z_i}|\Psi_0\rangle. \tag{5.16}$$

$$P_z^{(\mathrm{el})} = \frac{e}{2\pi L^2} \mathrm{Im} \log \langle \Psi_0 | \mathrm{e}^{-i\frac{2\pi}{L}\sum_i z_i} | \Psi_0 \rangle. \tag{5.17}$$

The proof of Eq. (5.17) is provided in the original paper [92], as well as in some review papers [3, 6, 7, 8].

Here we content ourselves to prove that Eq. (5.17) coincides with Eq. (5.12) in the special case where $|\Psi_0\rangle$ is the Slater determinant of the (doubly occupied) $\mathbf{k} = 0$ orbitals $|u_j\rangle = |\psi_j\rangle$. The key observation is that the many-body wavefunction

$$|\tilde{\Psi}_0\rangle = \mathrm{e}^{-i\frac{2\pi}{L}\sum_i z_i} |\Psi_0\rangle \tag{5.18}$$

is the Slater determinant built from the orbitals $|\tilde{u}_j\rangle = \mathrm{e}^{-i\frac{2\pi z}{L}} |u_j\rangle$, hence

$$\langle \Psi_0 | \mathrm{e}^{-i\frac{2\pi}{L}\sum_i z_i} | \Psi_0 \rangle = \langle \Psi_0 | \tilde{\Psi}_0 \rangle = (\,\det \langle u_j | \tilde{u}_{j'} \rangle\,)^2 = (\det\, S)^2, \tag{5.19}$$

where the second power owes to double orbital occupancy.

5.4.2 Kohn-Sham polarization vs. real polarization

All of the independent-electron formulas discussed in Sec. 5.3 are exact for noninteracting electrons, but the obvious aim is to implement them with KS orbitals, in a given density-functional theory (DFT) framework. Since macroscopic polarization applies to insulators only, we stress that we mean “KS insulator” throughout: that is, we assume that the KS spectrum is gapped. In the class of “simple” (i.e. computationally friendly) materials a genuine insulator is also a KS insulator, although pathological cases (computationally unfriendly) do exist.

Having specified this, the key issue is then: Does the KS polarization coincide with the physical many-body one? The answer is subtle, and is different whether one chooses either “open” boundary conditions, as appropriate for molecules and clusters, or periodic boundary conditions (Born-von Kármán), as invariably done in the present Notes.

Within open boundary conditions the KS orbitals vanish at infinity, as well as the charge density of the sample. $\mathbf{P}$ is then the first moment of the charge density, divided by the sample volume. The basic tenet of DFT is that the microscopic density of the fictitious noninteracting KS system coincides with the density of the interacting system. Therefore the exact $\mathbf{P}$ coincides by definition with the one obtained from the KS orbitals.

Matters are quite different within periodic boundary conditions: we have seen above that $\mathbf{P}$ is *not* a function of the charge density, hence the value of $\mathbf{P}$ obtained from the KS orbitals, in general, is not the correct many-body $\mathbf{P}$. This was first shown in 1995 by Gonze, Ghosez, and Godby [118], and later discussed by several authors. A complete account of the issue can be found in Ref. [6]. Here we just mention that the exact $\mathbf{P}$ is provided by Eq. (5.17), while the KS $\mathbf{P}$ is provided by Eq. (5.12), where the KS orbitals enter Eq. (5.11); both expressions are to be evaluated in the large-L limit. The two expressions are clearly different whenever the ground wave function is not a Slater determinant.

Therefore $\mathbf{P}$ cannot be exactly expressed within DFT, but the exact functional is obviously inaccessible, and even sometimes pathological. The practical issue is whether the current popular functionals provide an accurate approximation to the experimental values of $\mathbf{P}$ in a large class of materials.

A vast first-principle literature has accumulated over the years by either linear-response theory [119]—not reviewed here—or by the modern theory. The

errors are typically of the order of 10-20% on permittivity, and much less on most other properties (infrared spectra, piezoelectricity, ferroelectricity) for many different materials. It is unclear which part of the error is to be attributed to DFT per se, and which part is to be attributed to the *approximations* to DFT.

Chapter 6

Quantum metric and the theory of the insulating state

6.1 Nongeometrical theories of the insulating state

The standard textbook approach to the insulating state of matter is based on band theory. Following Bloch's theorem [120] in 1928, the main result is due to Wilson in 1931 [121]. The single-particle spectrum of a lattice-periodical Hamiltonian is in general gapped, and the electron count determines where the Fermi level lies. If it crosses a band one has a conductor: an applied electric field induces free acceleration of the electrons (at $T = 0$ in absence of dissipation). If the Fermi level lies instead in a gap, one has an insulator: in presence of a field the electronic system polarizes, but no steady-state current flows for $T \to 0$. This very successful theory explains the insulating/conducting behavior of most common materials across the periodic table, for which band structure calculations became soon available.

At the root of band theory are two basic assumptions: the electrons are noninteracting (in a mean-field sense), and the solid is crystalline. By the early 1960s, however, it became clear that there are solids where these two assumptions are very far from the truth, and where the insulating behavior is due to completely different mechanisms. The works of Mott in 1949 [122] and of Anderson in 1958 [123] opened new avenues in condensed matter physics. In the materials which we now call Mott insulators the insulating behavior is due to electron correlation [124], while in those called Anderson insulators it is due to lattice disorder [125].

In a milestone paper appeared in 1964 Kohn [126] provided a more comprehensive characterization of the insulating state of matter, which encompasses band insulators, Mott insulators, Anderson insulators, and eventually any kind of insulating material. According to Kohn, the electrons in the insulating state satisfy a many-electron localization condition [127]. This kind of localization must be defined in a subtle way given that, for instance, the Hamiltonian eigenstates in a band insulator are obviously *not* localized. According to the original Kohn's formulation, the insulating behavior arises whenever the ground-state wavefunction of an extended system breaks up into a sum of contributions which are localized in essentially disconnected regions of the many-electron configura-

tion space.

Kohn's theory remained little visited for many years [128] until the 1990s, when a breakthrough occurred in electronic structure theory: the modern theory of polarization. Inspired by the fact that electrical polarization discriminates *qualitatively* between insulators and metals, Resta and Sorella [129] in 1999 provided a definition of many-electron localization rather different from Kohn's, deeply rooted in the theory of polarization, and therefore based on geometrical concepts. Their program was completed soon after by Souza, Wilkens and Martin [130] (hereafter quoted as SWM), thus providing the foundations of the modern theory of the insulating state. An early review paper appeared in 2002 [4], and a recent one in 2011 [9]; the latter is at the root of the present Chapter.

6.2 Metric-curvature tensor

As in Sec. 5.4 we address an interacting N-electron system, whose most general Hamiltonian we write, in the Schrödinger representation and in Gaussian units, as

$$\hat{H}(\boldsymbol{\kappa}) = \frac{1}{2m_e} \sum_{i=1}^{N} |\mathbf{p}_i + \frac{e}{c}\mathbf{A}(\mathbf{r}_i) + \hbar\boldsymbol{\kappa}|^2 + \hat{V}. \tag{6.1}$$

Equation (6.1) is exact in the nonrelativistic, infinite-nuclear-mass limit. With respect to Eq. (5.14), the velocity is augmented with two terms: $\mathbf{A}(\mathbf{r})$ is a vector potential of magnetic origin, and $\boldsymbol{\kappa}$, having the dimensions of an inverse length, is a "flux" or "twist" of the same kind as the single-particle flux thoroughly discussed in Sec. 1.6.

In Sec. 5.4 we adopted periodic Born-von-Kàrmàn boundary conditions (PBCs) over a cubic supercell of side L, and the eigenfunctions $|\Psi_n\rangle$ were normalized in the hypercube of volume L^{3N}. In this Chapter, instead, it is expedient to consider in parallel both "open" boundary conditions (OBCs) and PBCs. The former are appropriate to molecular physics, and require that the many-electron wavefunction of a bound state is square-integrable over the whole coordinate space $\mathbb{R}^{3N}$. PBCs are instead appropriate for extended systems, either crystalline or disordered, either independent-electron or correlated. It is worth emphasizing that the choice of boundary conditions (either OBCs or PBCs) corresponds to choosing the Hilbert space, which in turn affects profoundly the geometrical properties discussed in Ch. 3.

6.2.1 Open boundary conditions

The case of OBCs is by far the simplest. We write for the sake of simplicity the ground state of Eq. (6.1) at $\boldsymbol{\kappa} = 0$ as $|\Psi_0\rangle \equiv |\Psi_0(0)\rangle$, and we define the many-body position operator as

$$\hat{\mathbf{r}} = \sum_{i=1}^{N} \mathbf{r}_i. \tag{6.2}$$

As in Sec. 1.6.2 the flux is easily "gauged away": the state $\mathrm{e}^{-i\boldsymbol{\kappa}\cdot\hat{\mathbf{r}}}|\Psi_0\rangle$ coincides with the ground eigenstate $|\Psi_0(\boldsymbol{\kappa})\rangle$ of the twisted Hamiltonian, Eq. (6.1). It

is legitimate to multiply this eigenvector by any $\boldsymbol{\kappa}$-dependent (and position-independent) phase factor; our choice is then

$$|\Psi_0(\boldsymbol{\kappa})\rangle = \mathrm{e}^{-i\boldsymbol{\kappa}\cdot(\hat{\mathbf{r}}-\mathbf{d})}|\Psi_0\rangle, \tag{6.3}$$

where $\mathbf{d} = \langle\Psi_0|\hat{\mathbf{r}}|\Psi_0\rangle$ is the electronic dipole of the molecular system. It follows that the $\boldsymbol{\kappa}$-derivative needed in our metric-curvature tensor, Eq. (3.24) is

$$|\nabla_{\boldsymbol{\kappa}}\Psi_0\rangle = -i(\hat{\mathbf{r}}-\mathbf{d})|\Psi_0\rangle = -i\hat{Q}(0)\,\hat{\mathbf{r}}|\Psi_0\rangle, \tag{6.4}$$

Exploiting the idempotency of the projector, the tensor $\mathcal{F}_{\alpha\beta}$, evaluated at $\boldsymbol{\kappa}=0$, is

$$\begin{aligned}\mathcal{F}_{\alpha\beta}(0) &= \langle\Psi_0|\hat{r}_\alpha\hat{Q}(0)\hat{r}_\beta|\Psi_0\rangle \\ &= \langle\Psi_0|\hat{r}_\alpha\hat{r}_\beta|\Psi_0\rangle - \langle\Psi_0|\hat{r}_\alpha|\Psi_0\rangle\langle\Psi_0|\hat{r}_\beta|\Psi_0\rangle,\end{aligned} \tag{6.5}$$

where $\hat{Q}$ is the projector over the unoccupied many-body eigenstates. $\mathcal{F}(0)$ clearly a real symmetric tensor, hence the Berry curvature vanishes within OBCs, and the metric $g_{\alpha\beta}(0)$ coincides with $\mathcal{F}_{\alpha\beta}(0)$. It will be shown that within PBCs, instead, the tensor $\mathcal{F}_{\alpha\beta}(0)$ may acquire a nonvanishing imaginary part.

Clearly, the metric tensor in Eq. (6.5) is the cumulant second moment of the position operator, or equivalently the ground state fluctuation of the dipole of the molecular system; this quantity is extensive (scales like N in macroscopically homogenous systems). We anticipate that $\mathcal{F}_{\alpha\beta}(0)/N$ discriminates, in the large N-limit, between insulators and metals.

6.2.2 Periodic boundary conditions

First of all, a key observation about the position operator is in order. The simple multiplicative operator $\hat{\mathbf{r}}$, as defined in Eq. (6.2), is "forbidden" within PBCs: in fact it maps any periodic wavefunction $|\Psi\rangle$ into the *nonperiodic* function $\hat{\mathbf{r}}|\Psi\rangle$. In other words, it maps a function which belongs to the Hilbert space to a function outside of it [92]. Incidentally, this is the main reason why the polarization problem has remained unsolved until the early 1990s. In the present context, formulae like Eq. (6.5) are ill defined and absurd within PBCs.

Within PBC the state $\mathrm{e}^{-i\boldsymbol{\kappa}\cdot\hat{\mathbf{r}}}|\Psi_0\rangle$ *is not* an eigenstate of Eq. (6.1), except for the discrete $\boldsymbol{\kappa}$ values

$$\boldsymbol{\kappa}_{m_1m_2m_3} = \frac{2\pi}{L}(m_1\mathbf{e}_1 + m_2\mathbf{e}_2 + m_3\mathbf{e}_3), \tag{6.6}$$

where $m_\alpha \in \mathbb{Z}$, and $\mathbf{e}_\alpha$ are the Cartesian versors. For fractional values of $\boldsymbol{\kappa}$—i.e. for values different from those in Eq. (6.6)—the $\boldsymbol{\kappa}$-dependence of the eigenvalues and eigenvectors of Eq. (6.1) is nontrivial. We have thoroughly discussed the single-particle version of this feature in Sec. 1.6.3.

If $|\Psi_0(\boldsymbol{\kappa})\rangle$ is the genuine ground eigenstate of Eq. (6.1) within PBCs, then the auxiliary function $|\tilde{\Psi}_0(\boldsymbol{\kappa})\rangle = \mathrm{e}^{i\boldsymbol{\kappa}\cdot\hat{\mathbf{r}}}|\Psi_0(\boldsymbol{\kappa})\rangle$ is a solution of $\hat{H}(0)$, but fulfills quasi-periodic boundary conditions: at any two opposite faces of the cube the wavefunction differs by a $\boldsymbol{\kappa}$-dependent phase factor. In other words the problem can be formulated in two equivalent ways: either the Hamiltonian is $\boldsymbol{\kappa}$-dependent, as in Eq. (6.1), and the boundary conditions are $\boldsymbol{\kappa}$-independent;

or the Hamiltonian is $\boldsymbol{\kappa}$-independent but the boundary conditions are "twisted" in a $\boldsymbol{\kappa}$-dependent way.

Within PBCs the tensor $\mathcal{F}_{\alpha\beta}(0)$ cannot be simplified—as e.g. in Eq. (6.5)—and must be therefore addressed in its original form by actually evaluating the $\boldsymbol{\kappa}$ derivatives:

$$\mathcal{F}_{\alpha\beta}(\boldsymbol{\kappa}) = \langle \partial_\alpha \Psi_0(\boldsymbol{\kappa}) | \hat{Q}(\boldsymbol{\kappa}) | \partial_\beta \Psi_0(\boldsymbol{\kappa}) \rangle. \tag{6.7}$$

As within OBCs, even within PBCs this tensor is extensive. In the thermodynamic limit (i.e $N \to \infty$, keeping N/L^3 constant), the well defined quantity is $\mathcal{F}_{\alpha\beta}(0)/N$. We also observe that for time-reversal symmetric systems $\mathcal{F}_{\alpha\beta}(0)$ is real symmetric, and coincides therefore with the metric tensor $g_{\alpha\beta}(0)$.

6.2.3 Sum over states again

We express $\langle r_\alpha r_\beta \rangle_{\rm c}$ using the sum-over-states formula, Eq. (3.28) :

$$\langle r_\alpha r_\beta \rangle_{\rm c} = \frac{1}{N} {\sum_{n\neq 0}}' \frac{\langle \Psi_0 | \partial_\alpha \hat{H}(0) | \Psi_n \rangle \langle \Psi_n | \partial_\beta \hat{H}(0) | \Psi_0 \rangle}{(E_0 - E_n)^2}. \tag{6.8}$$

The $\boldsymbol{\kappa}$-derivative of the Hamiltonian of Eq. (6.1) is

$$\nabla_{\boldsymbol{\kappa}} \hat{H}(0) = \frac{\hbar}{m_e} \sum_{i=1}^{N} \left[\mathbf{p}_i + \frac{e}{c} \mathbf{A}(\mathbf{r}_i) \right], \tag{6.9}$$

where the rhs is nothing else than $\hbar$ times the velocity operator $\hat{\mathbf{v}}$; Eq. (6.8) becomes then

$$\begin{aligned} \langle r_\alpha r_\beta \rangle_{\rm c} &= \frac{1}{\hbar^2 N} {\sum_{n\neq 0}}' \frac{\langle \Psi_0 | \hat{v}_\alpha | \Psi_n \rangle \langle \Psi_n | \hat{v}_\beta | \Psi_0 \rangle}{(E_0 - E_n)^2} \\ &= \frac{1}{N} {\sum_{n\neq 0}}' \frac{\langle \Psi_0 | \hat{v}_\alpha | \Psi_n \rangle \langle \Psi_n | \hat{v}_\beta | \Psi_0 \rangle}{\omega_{0n}^2}, \end{aligned} \tag{6.10}$$

where $\omega_{0n} = (E_n - E_0)/\hbar$.

The velocity operator is also commonly expressed as $\hat{\mathbf{v}} = i[\hat{H}(0), \hat{\mathbf{r}}]/\hbar$, but it is worth emphasizing that while the position $\hat{\mathbf{r}}$ is well defined within OBCs and ill defined within PBCs, the velocity $\hat{\mathbf{v}}$ is well defined in both cases.

The basic sum-over states formula, Eq. (6.10), applies therefore to both OBCs and PBCs. In general, it is not very useful on practical grounds, since it would require the evaluation of slowly convergent sums. Nonetheless, Eq. (6.10) is instead essential to gather understanding into the physical meaning of $\mathcal{F}_{\alpha\beta}$, as will be shown below.

6.3 Geometrical theory of the insulating state

6.3.1 Fundamentals

The modern formulation of the theory of the insulating state is based on a localization tensor (squared localization length in 1d), first introduced by Resta and Sorella in 1999 [129]. This work was followed soon afterwards by SWM

[130], where additional results are found, most notably relating localization to conductivity. Since then, most authors (including the present one) have adopted the notation $\langle r_\alpha r_\beta \rangle_\mathrm{c}$ for the localization tensor, where "c" stays for cumulant. The formulation within OBCs, using the language and the notations familiar in quantum chemistry, dates since 2006 [106].

The tensor $\langle r_\alpha r_\beta \rangle_\mathrm{c}$ has the dimensions of a squared length; it is an intensive quantity that characterizes the ground-state many-body wavefunction as a whole. Its key virtue is that it discriminates between insulators and metals: it is finite in the former case and divergent (in the large-system limit) in the latter. This is the main message of the present Chapter (and of the modern theory of the insulating state): we are going to prove it below, Sect. 6.4

Several expressions for the localization tensor have been given in the literature, all of them equivalent; here we define the localization tensor as the intensive quantity

$$\langle r_\alpha r_\beta \rangle_\mathrm{c} = \mathcal{F}_{\alpha\beta}(0)/N, \tag{6.11}$$

where the thermodynamic limit is understood. A glance at the OBCs expression, Eq. (6.5), shows the cumulant nature of $\mathcal{F}$ and explains the reason for the notation, which we adopt within PBCs as well.

Until 2005 the theory of the insulating state implicitly addressed time-reversal invariant systems only, where the $\mathcal{F}$ tensor is real symmetric, and coincides with the metric: in such systems therefore

$$\langle r_\alpha r_\beta \rangle_\mathrm{c} = g_{\alpha\beta}(0)/N. \tag{6.12}$$

It was found in 2005 [131] that—in absence of time-reversal symmetry and within PBCs—the tensor $\langle r_\alpha r_\beta \rangle_\mathrm{c}$ is naturally endowed with an antisymmetric imaginary part, whose physical meaning is also outstanding. This is discussed in Sects. **??** and **??**.

6.3.2 Linear response

So far, we have only discussed ground-state properties of our N-electron system. Suppose now that it is subject to a small time-dependent perturbation contributing to the Hamiltonian the term:

$$\delta\hat{H}(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} d\omega\; f(\omega)\frac{1}{2}(\hat{A}\mathrm{e}^{-i\omega t} + \hat{A}^\dagger \mathrm{e}^{i\omega t}), \tag{6.13}$$

where $\hat{A}$ determines the "shape" of the perturbation and f its amplitude. In order to get an Hermitian $\delta\hat{H}$, we assume $f(\omega) = f(-\omega)$. We wish to measure the response to such perturbation by means of the expectation value of some observable $\hat{B}$, i.e.:

$$\delta\langle\hat{B}\rangle = \langle\tilde{\Psi}|\hat{B}|\tilde{\Psi}\rangle - \langle\Psi_0|\hat{B}|\Psi_0\rangle, \tag{6.14}$$

where $\tilde{\Psi} = \Psi_0 + \delta\Psi(t)$ is the perturbed time-evolved ground state. If we limit ourselves to study terms which are linear in the response, it is enough to consider the single oscillatory perturbation:

$$\hat{H}'(\omega) = \frac{1}{2}(\hat{A}\mathrm{e}^{-i\omega t} + \hat{A}^\dagger \mathrm{e}^{i\omega t}), \tag{6.15}$$

whose response can be written, using the compact notations due to Zubarev [64, 132, 133], as:

$$\delta\langle\hat{B}\rangle = \frac{1}{2}(\,\langle\langle\hat{B}|\hat{A}\rangle\rangle_\omega \mathrm{e}^{-i\omega t} + \langle\langle\hat{B}|\hat{A}\rangle\rangle_{-\omega}\mathrm{e}^{i\omega t}\,). \tag{6.16}$$

The quantity $\langle\langle\hat{B}|\hat{A}\rangle\rangle_\omega$ is by definition the linear response induced by the perturbation $\hat{A}$ at frequency ω on the expectation value $\langle\hat{B}\rangle$. Straightforward first-order perturbation theory provides its explicit expression as:

$$\begin{aligned}\langle\langle\hat{B}|\hat{A}\rangle\rangle_\omega &= \frac{1}{\hbar}\lim_{\eta\to 0+}{\sum_{n\neq 0}}' \left(\frac{\langle\Psi_0|\hat{B}|\Psi_n\rangle\langle\Psi_n|\hat{A}|\Psi_0\rangle}{\omega-\omega_{0n}+i\eta}\right.\\ &- \left.\frac{\langle\Psi_0|\hat{A}|\Psi_n\rangle\langle\Psi_n|\hat{B}|\Psi_0\rangle}{\omega+\omega_{0n}+i\eta}\right), \end{aligned}\tag{6.17}$$

where ω_{0n} are the excitation frequencies of the unperturbed system, and the positive infinitesimal η ensures causality. Expressions of the kind of Eq. (6.17) go under the name of Kubo formulae.

6.3.3 Conductivity

The conductivity tensor $\sigma_{\alpha\beta}(\omega)$ measures the current linearly induced by an electric field: $j_\alpha = \sigma_{\alpha\beta}\mathcal{E}_\beta$. We therefore identify $\hat{A}$ with the potential of an electric field along β, i.e. $\hat{A} = e\mathcal{E}\hat{r_\beta}$, and $\hat{B}$ with the current operator $-e\hat{v}_\alpha/L^3$. An important detail must be stressed at this point. The macroscopic field inside the sample includes by definition screening effects due to the electronic system, while the perturbation $\delta\hat{H}$ entering Eq. (6.1)—via the $\hat{A}$ operator—is the "bare", or unscreened one. This point will be discussed below (Sect. 6.3.5); for the time being we simply identify screened and unscreened fields.

The Kubo formula for conductivity is therefore

$$\sigma_{\alpha\beta}(\omega) = -\frac{e^2}{L^3}\langle\langle\hat{v}_\alpha|\hat{r}_\beta\rangle\rangle_\omega; \tag{6.18}$$

this is correct within OBCs, but meaningless within PBCs, owing to the explicit presence of the position operator. This, however, makes no harm, since only its off-diagonal matrix elements are required: see Eq. (6.17). As usual, we may exploit the identity $\langle\Psi_0|\hat{\mathbf{r}}|\Psi_n\rangle = i\langle\Psi_0|\hat{\mathbf{v}}|\Psi_n\rangle/\omega_{0n}$. The Kubo formula becomes then

$$\begin{aligned}\sigma_{\alpha\beta}(\omega) &= \frac{ie^2}{\hbar L^3}\lim_{\eta\to 0+}{\sum_{n\neq 0}}' \frac{1}{\omega_{0n}}\left(\frac{\langle\Psi_0|\hat{v}_\alpha|\Psi_n\rangle\langle\Psi_n|\hat{v}_\beta|\Psi_0\rangle}{\omega-\omega_{0n}+i\eta}\right.\\ &+ \left.\frac{\langle\Psi_0|\hat{v}_\beta|\Psi_n\rangle\langle\Psi_n|\hat{v}_\alpha|\Psi_0\rangle}{\omega+\omega_{0n}+i\eta}\right). \end{aligned}\tag{6.19}$$

We introduce a compact notation for the real and imaginary parts of the numerators in Eq. (6.19), i.e.

$$\mathcal{R}_{n,\alpha\beta} = \mathrm{Re}\,\langle\Psi_0|\hat{v}_\alpha|\Psi_n\rangle\langle\Psi_n|\hat{v}_\beta|\Psi_0\rangle, \tag{6.20}$$

$$\mathcal{I}_{n,\alpha\beta} = \mathrm{Im}\,\langle\Psi_0|\hat{v}_\alpha|\Psi_n\rangle\langle\Psi_n|\hat{v}_\beta|\Psi_0\rangle, \tag{6.21}$$

which are symmetric and antisymmetric, respectively. Using then

$$\lim_{\mathcal{F}\to 0+} \frac{1}{x+i\eta} = \mathcal{P}\frac{1}{x} - i\pi\delta(x), \tag{6.22}$$

and omitting the principal part, we separate for $\omega > 0$ the symmetric and antisymmetric parts in the conductivity tensor as

$$\begin{aligned} \text{Re } \sigma^{(+)}_{\alpha\beta}(\omega) &= \frac{\pi e^2}{\hbar L^3} \sum_{n\neq 0}{}' \frac{\mathcal{R}_{n,\alpha\beta}}{\omega_{0n}} \delta(\omega - \omega_{0n}) \\ \text{Re } \sigma^{(-)}_{\alpha\beta}(\omega) &= = \frac{2e^2}{\hbar L^3} \sum_{n\neq 0}{}' \frac{\mathcal{I}_{n,\alpha\beta}}{\omega^2_{0n} - \omega^2}. \end{aligned} \tag{6.23}$$

6.3.4 Sum rules

At this point, we are ready to compare with the sum-over-states formulae for the tensor $\langle r_\alpha r_\beta \rangle_c$. In the present notations, we rewrite Eq. (6.10) as

$$\begin{aligned} \text{Re } \langle r_\alpha r_\beta \rangle_c &= \frac{1}{N} \sum_{n\neq 0}{}' \frac{\mathcal{R}_{n,\alpha\beta}}{\omega^2_{0n}} \\ \text{Im } \langle r_\alpha r_\beta \rangle_c &= \frac{1}{N} \sum_{n\neq 0}{}' \frac{\mathcal{I}_{n,\alpha\beta}}{\omega^2_{0n}}. \end{aligned} \tag{6.24}$$

A glance at Eq. (6.23) shows that

$$\text{Re } \langle r_\alpha r_\beta \rangle_c = \frac{\hbar L^3}{\pi e^2 N} \int_0^\infty \frac{d\omega}{\omega} \text{ Re } \sigma^{(+)}_{\alpha\beta}(\omega) \tag{6.25}$$

$$\text{Im } \langle r_\alpha r_\beta \rangle_c = \frac{\hbar L^3}{2e^2 N} \text{Re } \sigma^{(-)}_{\alpha\beta}(0). \tag{6.26}$$

Eq. (6.25) has been arrived at by SWM in 2000 [130], and Eq. (6.26) by Resta in 2005 [131]. First of all, these identities show that $\langle r_\alpha r_\beta \rangle_c$, defined here as a basic geometric feature, is indeed a measurable quantity (whenever it does not diverge).

As emphasized throughout this work $\langle r_\alpha r_\beta \rangle_c$ is a ground-state property, while the rhs of Eqs. (6.25) and (6.26) are properties of the system *excitations*, owing to the Kubo formula. Indeed, both Eqs. (6.25) and (6.26) look like the zero-temperature limit of a fluctuation-dissipation theorem, several forms of which are known in statistical physics [134, 135]: in the lhs we have a ground-state fluctuation—see in particular Eq. (6.5)—while the ingredient of the rhs is conductivity (dissipation).

6.3.5 Screened vs. unscreened field

The Kubo formula for conductivity has been obtained identifying the $\hat{A}$ operator with $\mathcal{E}\hat{r_\beta}$, where $\mathcal{E}$ is the macroscopic field inside the sample. In general this is not quite correct, since instead $\hat{A} = e\mathcal{E}_0\hat{r}_\beta$, where $\mathcal{E}_0$ is the "bare" field, i.e. the field that would be present inside the sample *in absence of screening.* The latter originates from the two-body (electron-electron) terms in the potential $\hat{V}$

entering Schrödinger equation. The relationship between $\mathcal{E}$ and $\mathcal{E}_0$ is not a bulk property, and depends on the shape of the sample. Alternatively, it depends on the boundary conditions assumed for integrating Poisson equation: we refer to Ref. [8] for a thorough discussion. Whenever $\mathcal{E} \neq \mathcal{E}_0$, the sum rule in Eq. (6.25) must be modified.

The ground-state fluctuations, as e.g. Eq. (6.5), are fluctuations of the macroscopic polarization, which in a finite sample induce a surface charge at the boundary. This in turn generates a depolarizing field, which counteracts polarization. Therefore the localization tensor $\langle r_\alpha r_\beta \rangle_\mathrm{c}$ depends on the sample shape, or equivalently on the boundary conditions assumed when taking the thermodynamic limit. The choice of PBCs in Eq. (6.1), however, implies $\mathcal{E} = \mathcal{E}_0$ [136]; hence Eqs. (6.25) and (6.26) are correct as they stand within PBCs. This no longer holds within OBCs: in this case Eq. (6.25) needs to be modified, while Im $\langle r_\alpha r_\beta \rangle_\mathrm{c} = 0$.

Ideally the equality $\mathcal{E} = \mathcal{E}_0$ corresponds to choosing a sample in the form of a slab, and to addressing the component of the fluctuation tensor $\langle r_\alpha r_\beta \rangle_\mathrm{c}$ parallel to the slab [8]; the thermodynamic limit amounts then to the infinite slab thickness. Owing to the long range of Coulomb interaction the order of the limits (first a slab, then its infinite thickness) is crucial. For instance, if the limit is taken instead by considering spherical clusters of increasing radius, the SWM fluctuation-dissipation sum rule, Eq. (6.25), assumes a different form: this is discussed in Refs. [106, 136]. The explicit form of the generalized sum rule is given therein.

Last but not least, the effect leading to $\mathcal{E} \neq \mathcal{E}_0$ within OBCs is a pure correlation effect. It originates from explicitly correlated wavefunctions, and does not occur within mean-field theories (Hartree-Fock and Kohn-Sham) [106, 136]. Within such theories, therefore, the sum rules hold in the simple form of Eqs. (6.25) and (6.26); the conductivity therein is the independent-particle conductivity ("uncoupled" response in quantum-chemistry jargon).

6.4 Localization in the insulating state

The basic tenet of the modern theory of the insulating state is that the localization tensor $\langle r_\alpha r_\beta \rangle_\mathrm{c}$ is the ground-state property which sharply discriminates—in the spirit of Kohn's seminal work [126, 127]—between insulators and metals. The real part of $\langle r_\alpha r_\beta \rangle_\mathrm{c}$ remains finite in the thermodynamic limit in any insulator, while it diverges in any metal.

The theory is very general, and has found applications to various different kinds of insulators: band insulators [137, 138, 139, 140]; correlated (i.e. Mott) insulators, either by means of Hubbard-like model Hamiltonians [129, 141] or realistic ones [142, 143]; and Anderson insulators [144]. Some of these applications are reviewed in Sect. 6.5. The cases of Chern insulators, quantum Hall insulators, and topological insulators are discussed in Chap. **??**.

The ultimate proof of the key property of Re $\langle r_\alpha r_\beta \rangle_\mathrm{c}$ is based on the SWM sum rule, Eq. (6.25). Since the tensor is real symmetric, it is enough to consider the diagonal elements (over its principal axes)

$$\mathrm{Re}\ \langle r_\alpha r_\alpha \rangle_\mathrm{c} = \frac{\hbar L^3}{\pi e^2 N} \int_0^\infty \frac{d\omega}{\omega}\ \mathrm{Re}\ \sigma_{\alpha\alpha}(\omega). \tag{6.27}$$

The f-sum rule yields

$$\int_0^\infty d\omega \text{ Re } \sigma_{\alpha\alpha}(\omega) = \frac{\omega_{\rm p}^2}{8} = \frac{\pi e^2 N}{2m_e L^3}, \tag{6.28}$$

where $\omega_{\rm p}$ is the plasma frequency. Therefore the integral in Eq. (6.27) always converges at ∞; its convergence/divergence is dominated by the small-ω behavior of Re $\sigma_{\alpha\alpha}(\omega)$.

Suppose first that the spectrum is gapped, i.e. the spacing between the ground state and the first excited state stays finite in the thermodynamic limit. If the gap is $E_{\rm g}$ the conductivity vanishes for $\omega < E_{\rm g}/\hbar$, and Eq. (6.28) yields

$$\begin{aligned} \text{Re } \langle r_\alpha r_\alpha \rangle_{\rm c} &= \frac{\hbar L^3}{\pi e^2 N} \int_{E_{\rm g}/\hbar}^\infty \frac{d\omega}{\omega} \text{ Re } \sigma_{\alpha\alpha}(\omega) \\ &< \frac{\hbar^2 L^3}{\pi e^2 N E_{\rm g}} \int_0^\infty d\omega \text{ Re } \sigma_{\alpha\alpha}(\omega) \\ &= \frac{\hbar^2}{2m_e E_{\rm g}}. \end{aligned} \tag{6.29}$$

This inequality is due to SWM and clearly proves that Re $\langle r_\alpha r_\beta \rangle_{\rm c}$ is finite in any gapped insulator, as e.g. band insulators (considered in more detail in Sect. 6.4.2).

The main message of Kohn's 1964 paper, however, is that "insulating characteristics are a strict consequence of electronic localization (in an appropriate sense) and do not require an energy gap". For any gapless material, the small-ω behavior of Re $\sigma_{\alpha\alpha}(\omega)$ is the result of a competition between numerators and denominators in the Kubo formula, Eq. (6.19). Since we aim at a continuous function of ω, the singularities in Eq. (6.23) must be smoothed: this can be done by keeping the "dissipation" η finite while performing the thermodynamic limit first [145]. For a band metal the localization tensor diverges (see below). According to SWM, a gapless material is insulating whenever Re $\sigma_{\alpha\alpha}(\omega) \to 0$ like a positive power of ω, and metallic otherwise. The only example of gapless insulator considered so far is a model Anderson insulator in 1d [144]. Simulations prove indeed that $\langle x^2 \rangle_{\rm c}$ is finite therein (Sect. 6.5.4).

6.4.1 Independent electrons

For noninteracting electrons the potential $\hat{V}$ in Eq. (6.1) is the sum of identical one-body terms: $\hat{V} = \sum_{i=1}^N V(\mathbf{r}_i)$. The many-electron Hamiltonian is separable and the exact ground state $|\Psi_0(\boldsymbol{\kappa})\rangle$ is a Slater determinant of one-particle orbitals (doubly occupied in the singlet case). At a mean-field level, the one-body potential $V(\mathbf{r})$ includes electron-electron interaction in a selfconsistent way. In the Hartree-Fock (HF) framework the Slater determinant is regarded as an approximate many-electron wavefunction. Instead, in the density-functional framework the orbitals—called Kohn-Sham (KS) orbitals—are auxiliary quantities, individually devoid of physical meaning. In particular, their Slater determinant *does not* coincide with the many-electron wavefunction $|\Psi_0(\boldsymbol{\kappa})\rangle$, as a matter of principle. Therefore the exact localization tensor does not coincide, at least in principle, with the one obtained from the Slater determinant of KS

orbitals; the issue is similar to the one discussed above for macroscopic polarization $\mathbf{P}$ (Sec. 5.4.2). We stress, however, a difference: the exact $\mathbf{P}$ coincides with the KS $\mathbf{P}$ within OBCs, and the problem occurs only within PBCs. Instead, the exact localization tensor differs in principle from the KS one within both OBCs and PBCs.

So much for the matters of principle. On practical grounds such difference is routinely disregarded (e.g. when dealing with polarization, magnetization [5, 6, 8], and more), given that it is not at all clear what is the relative importance of this "intrinsic" error, compared with the errors due to the choice of the functional itself. We therefore address here either the HF or the KS wavefunction $|\Psi_0(\boldsymbol{\kappa})\rangle$, having the form of a Slater determinant.

Whenever the wavefunction is a Slater determinant, all ground-state properties can be explicitly cast in terms of the one-body density matrix

$$\rho(\mathbf{r},\mathbf{r}') = 2P(\mathbf{r},\mathbf{r}') = 2\sum_{j=1}^{N/2} \varphi_j(\mathbf{r})\varphi_j^*(\mathbf{r}'), \tag{6.30}$$

where a singlet ground state is assumed, and $\varphi_j(\mathbf{r})$ are the occupied one-particle orbitals (either HF or KS); $P(\mathbf{r},\mathbf{r}')$ is the projector over the occupied manifold.

The expression for the localization tensor is easily found within OBCs starting from Eq. (6.5) [106]:

$$\langle r_\alpha r_\beta \rangle_{\rm c} = \frac{1}{N} \int d\mathbf{r} d\mathbf{r}' \; (\mathbf{r}-\mathbf{r}')_\alpha (\mathbf{r}-\mathbf{r}')_\beta \, |P(\mathbf{r},\mathbf{r}')|^2. \tag{6.31}$$

If we define the complementary projector

$$Q(\mathbf{r},\mathbf{r}') = \delta(\mathbf{r}-\mathbf{r}') - P(\mathbf{r},\mathbf{r}'), \tag{6.32}$$

an equivalent expression is [137]

$$\langle r_\alpha r_\beta \rangle_{\rm c} = \frac{2}{N} \mathrm{Tr}\ \{ r_\alpha P r_\beta Q \}, \tag{6.33}$$

where "Tr" is the trace over the single-particle Hilbert space (not on Cartesian indices).

If we consider a cluster, cut out of a crystalline solid, Eq. (6.31) becomes in the large-N limit

$$\langle r_\alpha r_\beta \rangle_{\rm c} = \frac{1}{N_{\rm c}} \int_{\rm cell} d\mathbf{r} \int_{\text{all space}} d\mathbf{r}' \; (\mathbf{r}-\mathbf{r}')_\alpha (\mathbf{r}-\mathbf{r}')_\beta \, |P(\mathbf{r},\mathbf{r}')|^2, \tag{6.34}$$

where $N_{\rm c}$ is the number of electrons per crystal cell. According to the discussion in Sec. 6.3.5, we need not to worry about shape issues in taking the limit; we also notice that the density matrix, Eq. (6.30), is independent of the boundary conditions (either OBCs or PBCs) in the large-N limit.

6.4.2 Band insulators and band metals

As observed, Eq. (6.34) holds for a crystalline solids. Therefore the inner integral on the rhs must converge for a band insulator, and must diverge for a band metal. This is confirmed by the well known fact that the asymptotic behavior of P is

qualitatively different in insulators and in metals. In the former materials, in fact, $P(\mathbf{r}, \mathbf{r}')$ decays exponentially [146, 147, 148, 149] for large values of $\mathbf{r} - \mathbf{r}'$: therefore the integral converges and the localization tensor is finite. In conducting materials, instead, $P(\mathbf{r}, \mathbf{r}')$ decays only polynominially, and the inner integral diverges. This divergence can be explicitly verified for the simplest conductor of all, namely, the noninteracting electron gas, whose density matrix is exactly known in analytic form [137, 150]. Therefore the localization tensor, when expressed in the form of Eq. (6.34), measures in a perspicuous way the "nearsightedness"[151] of the electron distribution. Such measure is *qualitatively* different in insulators and in metals.

The one-particle orbitals (either HF or KS) in a crystalline solid have the Bloch form. We therefore may wish to replace the orbitals in the expression for $P(\mathbf{r}, \mathbf{r}')$, Eq. (6.30),

$$\varphi_j(\mathbf{r}) \rightarrow \psi_{j\mathbf{k}}(\mathbf{r}) = \mathrm{e}^{i\mathbf{k}\cdot\mathbf{r}} u_{j\mathbf{k}}(\mathbf{r}), \tag{6.35}$$

where j is the band index and $\mathbf{k}$ is the Bloch vector. We stress that PBCs are at the very root of Bloch theorem. If the orbitals are normalized to one over the crystal cell of volume V_{cell}, the ground-state projector in insulating crystals is

$$\begin{aligned} P(\mathbf{r}, \mathbf{r}') &= \frac{V_{\text{c}}}{(2\pi)^3} \sum_{j=1}^{n} \int_{\text{BZ}} d\mathbf{k}\; \psi_{j\mathbf{k}}(\mathbf{r}) \psi^*_{j\mathbf{k}}(\mathbf{r}') \qquad (6.36) \\ &= \frac{V_{\text{cell}}}{(2\pi)^3} \sum_{j=1}^{n} \int_{\text{BZ}} d\mathbf{k}\; \mathrm{e}^{i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}')} u_{j\mathbf{k}}(\mathbf{r}) u^*_{j\mathbf{k}}(\mathbf{r}'), \end{aligned}$$

where $n = N_{\text{c}}/2$ is the number of occupied bands, and the integral is taken over the Brillouin zone (equivalently, over the reciprocal cell).

Using Eq. (6.36), the localization tensor in a band insulator becomes (for double occupancy) proportional to BZ integral of the Bloch metric-curvature tensor, Eq. (3.47), i.e.

$$\begin{aligned} \langle r_\alpha r_\beta \rangle_{\text{c}} &= \frac{V_{\text{cell}}}{(2\pi)^3\, n} \int_{\text{BZ}} d\mathbf{k}\; \mathcal{F}_{\alpha\beta}(\mathbf{k}) \\ \mathcal{F}_{\alpha\beta}(\mathbf{k}) &= \sum_{j=1}^{n} \langle \partial_\alpha u_{j\mathbf{k}} | \partial_\beta u_{j\mathbf{k}} \rangle - \sum_{jj'=1}^{n} \langle \partial_\alpha u_{j\mathbf{k}} | u_{j'\mathbf{k}} \rangle \langle u_{j'\mathbf{k}} | \partial_\beta u_{j\mathbf{k}} \rangle. \qquad (6.37) \end{aligned}$$

The proof is given in Refs. [4, 137], and will not be repeated here.

Within OBCs the localization tensor is always real: this is perspicuous in Eqs. (6.31) and (6.34). Instead the imaginary part of the PBC $\langle r_\alpha r_\beta \rangle_{\text{c}}$, Eq. (6.37), is the BZ integral of the Bloch Berry curvature, discussed below, Sec. xxxx.

6.4.3 Wannier functions

Expressions similar to Eq. (6.37) also enter the Marzari-Vanderbilt theory of maximally localized Wannier functions [103]; the relationship is

$$\sum_{\alpha=1}^{d} \langle r_\alpha r_\alpha \rangle_{\text{c}} = \frac{1}{n} \Omega_{\text{I}}. \tag{6.38}$$

Here $\Omega_{\rm I}$ indicates the gauge-invariant part of the quadratic spread of the Wannier functions, as in Ref. [103] and in the subsequent literature. The trace in Eq. (6.38) is a lower bound (and *not* a minimum in dimension $\mathrm{d} > 1$) for the spherical second (cumulant) moment—a.k.a. quadratic spread—of the Wannier functions, averaged over the sample.

We stress that Eqs. (6.36) and (6.37) make sense only insofar the Fermi level falls in a gap, in which case $\langle r_\alpha r_\beta \rangle_{\rm c}$ is always finite. If we vary the Hamiltonian continuously, allowing the gap to close, then $\langle r_\alpha r_\beta \rangle_{\rm c}$ diverges; the quadratic spread of the Wannier functions diverges as well [106].

6.5 Localization in different kinds of insulators

6.5.1 Small molecules

The modern theory of the insulating state clearly addresses extended systems, i.e. the $N \to \infty$ limit; indeed it makes little sense to ask whether a small molecule is insulating or conducting. Nonetheless the concepts of localized/delocalized electronic states is of the utmost importance in quantum chemistry as well, notably in relationship to aromaticity.

The tensor $\langle r_\alpha r_\beta \rangle_{\rm c}$ within OBCs is always real symmetric. If the ground-state wavefunction is a Slater determinant, then the trace of the tensor at finite N has the meaning of a lower bound for the quadratic spread of the Boys localized orbitals, averaged over all the occupied orbitals [106].

The small-N version of the main concepts of the present review (in their OBCs flavour [106]) has been adopted in quantum chemistry by Àngyàn [152, 153]. Besides providing HF calculations of $\langle r_\alpha r_\beta \rangle_{\rm c}$ for a sample of small molecules, Àngyàn even provides *experimental values* drawn from compilations of the dipole oscillation-strength distributions: basically, from Eq. (6.25).

6.5.2 Band insulators

The theory warrants that the localization tensor is finite in any insulator. However, quantitative calculations for both model tight-binding Hamiltonians and realistic solids within density-functional theory have been used to illustrate the theory and to identify trends. For instance one expects much smaller diagonal elements $\langle r_\alpha r_\alpha \rangle_{\rm c}$ for strong (i.e. large-gap) insulators than for weak (small-gap) ones. This is also suggested by the SWM inequality, Eq. (6.29).

Let us start with a simple tight-binding (a.k.a. Hückel) Hamiltonian in 1d:

$$\hat{H} = \sum_{j\sigma} [\, (-1)^j \Delta\, c^\dagger_{j\sigma} c_{j\sigma} - t(c^\dagger_{j\sigma} c_{j+1\sigma} + \text{H.c.})\,] \tag{6.39}$$

where $t > 0$ is the first neighbor hopping ($\beta = -t$ in most chemistry literature) and H.c. stays for Hermitian conjugate. This toy model schematizes a binary ionic crystal; the band structure is

$$\epsilon(q) = \pm\sqrt{\Delta^2 + 4t^2 \cos^2 qa/2}, \tag{6.40}$$

where a is the lattice constant and q is the Bloch vector. The gap is equal to 2Δ; at half filling the system is always insulating except for $\Delta = 0$. The squared

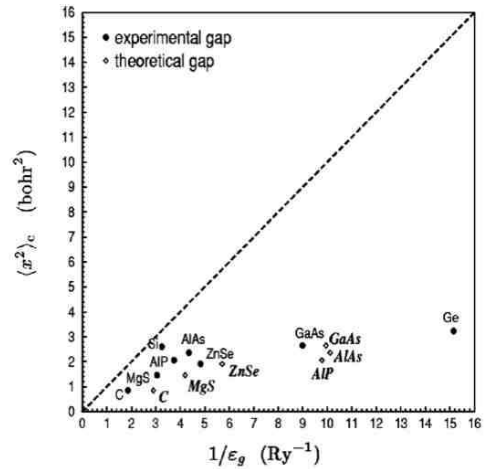

Figure 6.1: Diagonal element of the KS localization tensor vs. the inverse direct gap (theoretical and experimental), for several elemental and binary semiconductors (from Ref. [137]) The points corresponding to Si and Ge with the theoretical gaps are out of scale. From Ref. [137].

localization length (within OBCs) is the tight-binding version of Eq. (6.31), i.e.

$$\langle x^2 \rangle_c = \frac{a^2}{4N} \sum_{j,j'=1}^{N} P_{jj'}^2 (j - j')^2. \tag{6.41}$$

This is a monothonical function of t/Δ; it is easily verified that it vanishes in the extreme ionic case ($t = 0$). In the metallic case ($\Delta = 0$) the ground-state projector has a simple analytical form:

$$\begin{aligned} P_{jj} &= \frac{1}{2}; \quad P_{jj'} = 0 \text{ for even } |j' - j| = 2s, \\ P_{jj'} &= \frac{(-1)^s}{\pi(2s+1)} \text{ for odd } |j' - j| = 2s + 1, \end{aligned} \tag{6.42}$$

which clearly implies divergence of Eq. (6.41). At any finite N within OBCs Eq. (6.41) leads to a finite $\langle x^2 \rangle_c$ value; however Eq. (6.42) suggests that in the metallic case $\langle x^2 \rangle_c$ diverges linearly with N. This has been verified by actual simulations, even when $\Delta \neq 0$ but the Fermi level is not in the gap [144].

Other simulations [140] have addressed dimerized chains, i.e. $\Delta = 0$ but alternant hoppings in Eq. (6.39). While nothing relevant occurs within PBCs, partly filled end states within OBCs at some fillings are at the root of some noticeable features.

The first ab-initio study (in 2001) addressed several elemental and binary cubic semiconductors at the KS level [137]. The tensor is real and isotropic. The computed $\langle x^2 \rangle_c$ (Fig. 6.1) is smaller than 3 bohr^2 in all the materials studied: the ground many-body wavefunction is therefore very localized in this class of materials. The SWM inequality was also checked, and found to be well verified using both the theoretical KS gap and the experimental one (the latter is typically larger).

Other studies have addressed the ferroelectric perovskites in their different (cubic and noncubic) structures [138], and some model Hamiltonians in 1d and 2d [139].

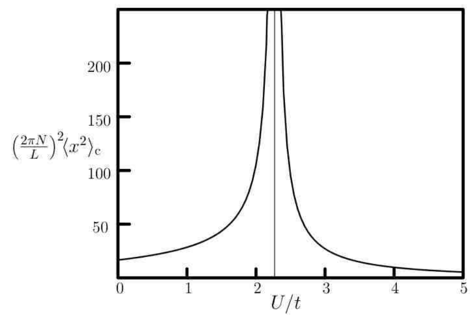

Figure 6.2: Squared localization length for the Hamiltonian in Eq. (6.43) at half filling for $t/\Delta = 1.75$. The system undergoes a quantum phase transition from band-like insulator to Mott-like insulator at $U/t = 2.27$. From Ref. [129].

6.5.3 Correlated (Mott) insulators

Starting from the noninteracting Hamiltonian of Eq. (6.39) and augmenting it with an on-site repulsive term we get the two-band Hubbard model

$$\hat{H} = \sum_{j\sigma} [(-1)^j \Delta\, c^\dagger_{j\sigma} c_{j\sigma} - t(c^\dagger_{j\sigma} c_{j+1\sigma} + \text{H.c.})] + U \sum_j n_{j\uparrow} n_{j\downarrow}. \tag{6.43}$$

The explicitly correlated ground-state wavefunction has been found by exact diagonalization [129], and the corresponding $\langle x^2 \rangle_c$ has been computed as a function of U for fixed $t/\Delta = 1.75$. The results are shown in Fig. 6.2 in dimensionless units; it turns out that there is only one singular point $U = 2.27t$, where $\langle x^2 \rangle_c$ diverges. Indeed, it has been verified that at such value the ground-state becomes degenerate with the first excited singlet state, i.e. the system is metallic. The singular point is the fingerprint of a quantum phase transition: on the left we have a band-like insulator, and on the right a Mott-like insulator. The two insulating states are *qualitatively* different; by adopting the modern jargon, nowadays we could say that they are *topologically* distinct. The static ionic charges (on anion and cation) are continuous across the transition, while the dynamical (Born) effective charge on a given site changes sign [154]. Other studies of the localization tensor within the same Hubbard model can be found in Ref. [155].

The transition from a band metal to a Mott insulator has been studied in a model linear chain of Li atoms by Vetere *et al.* [142]. At a mean-field level the infinite chain is obviously metallic at any lattice constant a, since there is one valence electron per cell. However the mean-field description becomes inadequate at large a, where the electrons localize and the system becomes a Mott insulator. If electron correlation is properly accounted for at any a, the system undergoes a sharp metal-insulator transition at a critical a.

The calculations addressed linear Li_N systems (N up to 8) within OBCs, where the finite size prevents a sharp transition; the tradeoff is that full configuration interaction was affordable with 6 atomic orbitals per site (yielding more than 10^9 symmetry-adapted Slater determinants). The wavefunction of Vetere *et al.* is therefore exempt from any bias insofar as the treatment of correlation is concerned, although its quality is determined by the basis set. A study of the longitudinal component $\langle x^2 \rangle_c$ of the localization tensor indicates rather clearly the occurrence of the metal-insulator transition at $a \simeq 7$ bohr; other indicators give concordant results [142]. For comparison, the nearest-neighbour distance in 3d metallic lithium is 5.73 bohr.

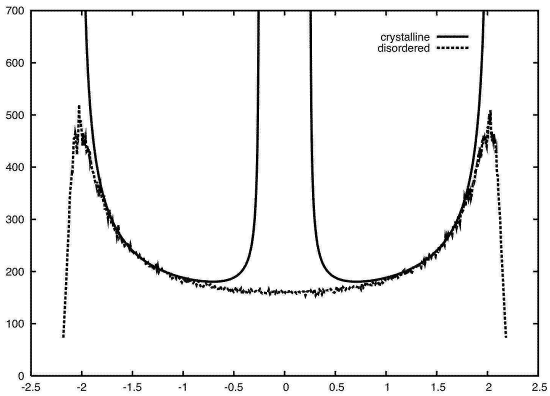

Figure 6.3: Density of states (arbitrary units) for a model binary alloy in 1d. The crystalline (band) case corresponds to the Hamiltonian of Eq. (6.39) with $\Delta = 0.25$ and $t = 1$. The disordered (Anderson) case corresponds to a random choice of the anion/cation distribution.

More recently, Stella et al. [143] have performed variational quantum Monte Carlo studies of hydrogen chains, up to 66 atoms, within PBCs. The crossover between the weakly correlated (band) metallic regime—at short distances—and the strongly correlated (Mott) insulating regime—at large distances—has been thoroughly investigated by means of $\langle x^2 \rangle_c$. The Mott transition occurs at $a \simeq 3.5$ bohr.

6.5.4 Disordered (Anderson) insulators

We start from the same Hamiltonian as in Eq. (6.39), and we replace the ordered string $(-1)^j$ by a random string of ± 1, chosen with equal (and uncorrelated) probability. This system models a random binary alloy at 50% concentration. It is well known both from analytical arguments and actual simulations that its spectrum is gapless [125, 156]. The density of states for both the ordered and disordered systems are shown in Fig. 6.3, and confirm the expected features. The band structure of Eq. (6.40) yields obviously a gapped density of states; at the band edges it shows van Hove singularities, which in 1d have the character of $1/\sqrt{\epsilon}$ divergences. As discussed above, the system is insulating at half filling and conducting otherwise. The disordered system, instead, is gapless and nonetheless insulating at any filling. In fact, this model Hamiltonian describes a paradigmatic Anderson insulator in 1d.

The conventional theory of transport focusses on the nature of the one-particle orbitals at the Fermi level; in Anderson insulators these are localized, thus forbidding steady state currents [123]. More than fifty years of literature have been devoted to investigate Anderson insulators under the most diverse aspects. [125, 156, 157, 158].

At variance with such wisdom, a recent work has addressed this paradigmatic Anderson insulator from the nonconventional viewpoint of the modern theory of the insulating state [144]. In the spirit of Kohn's theory the invidual Hamiltonian eigenstates become apparently irrelevant, while the focus is on the many-electron ground state as a whole. The squared localization length $\langle x^2 \rangle_c$ has been computed within OBCs from Eq. (6.41), and found to be finite, as expected. Nonetheless its value is about 20 times larger than the one for the band insulator, at the same value of the parameters (i.e. $\Delta = 0.25, t = 1$). This reflects the fact that the scattering mechanisms are profoundly different: inco-

herent (Anderson) versus coherent (band). In the latter case, the Hamiltonian eigenstates are individually conducting but "locked" by the Pauli principle if the Fermi level lies in the gap.

Bibliography

[1] R. Resta, Rev. Mod. Phys. **66**, 899 (1994).

[2] R. Resta, in: *Quantum-Mechanical Ab-initio Calculation of the Properties of Crystalline Materials*, Lecture Notes in Chemistry, Vol. **67**, edited by C. Pisani (Springer, Berlin, 1996), p. 273.

[3] R. Resta, J. Phys.: Condens. Matter **12**, R107 (2000).

[4] R. Resta, J. Phys.: Condens. Matter **14**, R625 (2002).

[5] D. Vanderbilt and R. Resta, in: *Conceptual foundations of materials: A standard model for ground- and excited-state properties*, S.G. Louie and M.L. Cohen, eds. (Elsevier, 2006), p. 139.

[6] R. Resta and D. Vanderbilt, in: *Physics of Ferroelectrics: a Modern Perspective*, Topics in Applied Physics Vol. **105**, Ch. H. Ahn, K. M. Rabe, and J.-M. Triscone, eds. (Springer-Verlag, 2007), p. 31.

[7] R. Resta, J Phys.: Conference Series **117**, 012024 (2008).

[8] R. Resta, J. Phys.: Condens. Matter **22** 123201 (2010).

[9] R. Resta, Eur. Phys. J. B **79**, 121 (2011).

[10] M. Z. Hasan and C. L. Kane, Rev. Mod. Phys. **82**, 3045 (2010).

[11] N. Nagaosa, J. Sinova, S. Onoda, A. H. MacDonald, and N. P. Ong, Rev. Mod. Phys. **82**, 1539 (2010).

[12] D. Xiao, M.-C. Chang, and Q. Niu, Rev. Mod. Phys. **82**, 1959 (2010).

[13] E. Prodan, J. Phys. A **44**, 113001 (2011).

[14] X.-L. Qi and S.-C. Zhang, Rev. Mod Phys. **83**, 1057 (2011).

[15] M. Z. Hasan and J. E. Moore, Annu. Rev. Condens. Matter Phys. **2**, 55 (2011).

[16] J. Maciejko, T. L. Hughes, and S.-C. Zhang, Annu. Rev. Condens. Matter Phys. **2**, 31 (2011).

[17] T. Thonhauser, Int. J. Mod. Phys. B **25**, 1429 (2011).

[18] N. Marzari, A. A. Mostofi, J. R. Yates, I. Souza, and D. Vanderbilt, `http://arxiv.org/abs/1112.5411`; Rev. Mod. Phys., in press.

[19] R. Resta, *Berry's Phase and Geometric Quantum Distance: Macroscopic Polarization and Electron Localization*, Lecture Notes for the "Troisième Cycle de la Physique en Suisse Romande" (Lausanne, 2000). `http://www-dft.ts.infn.it/~resta/publ/notes2000.ps`.

[20] Y. Aharonov and D. Bohm, Phys. Rev. **115**, 485 (1959); reprinted in *Geometric Phases in Physics*, edited by A. Shapere and F. Wilczek (World Scientific, Singapore, 1989), p.104.

[21] R. P. Feynman, R. B. Leighton, and M. Sands, *The Feynman Lectures in Physics, Vol. 2* (Addison Wesley, Reading, 1964), Sect. 15-4.

[22] R. G. Chambers, Phys. Rev. Lett. **5**, 3 (1960).

[23] M. Peshkin and A. Tonomura, *The Aharonov–Bohm Effect* (Springer, Berlin, 1989).

[24] `http://en.wikipedia.org/wiki/SQUID`.

[25] P. Bocchieri and A. Loinger, Nuovo Cimento **47**, 475 (1978).

[26] `http://www.phy.bris.ac.uk/people/berry_mv/index.html`.

[27] M. V. Berry, Proc. Roy. Soc. Lond. A **392**, 45 (1984).

[28] H. C. Longuet-Higgins, U. Öpik, M. H. L. Pryce, and R. A. Sack, Proc. Roy. Soc. A **244**, 1 (1958).

[29] G. Herzberg and H. C. Longuet-Higgins, Discuss. Faraday Soc. **35**, 77 (1963).

[30] C. A. Mead and D. G. Truhlar, J. Chem. Phys. **70**, 2284 (1979).

[31] C. A. Mead, Chemical Physics **49**, 23 (1980).

[32] V. Heine, Phys. Rev.**145**,593 (1966).

[33] J. A. Appelbaum and D. R. Hamann, Phys. Rev. B **10**, 4973 (1974).

[34] L. Kleinman, Phys. Rev. B **11**, 858 (1975).

[35] F. Claro, Phys. Rev. B **17**, 699 (1977).

[36] J. A. Appelbaum, G. A. Baraff, and D. R. Hamann, Phys. Rev. B **14**, 1623 (1976).

[37] K. von Klitzing, G. Dorda, and M. Pepper, Phys. Rev. Lett. **45**, 494 (1980).

[38] T. Ando, J. Phys. Soc. Jpn. **37**, 622 (1974).

[39] R. B. Laughlin, Phys. Rev. B **23**, 5632 (1981).

[40] R. E. Prange, S. M. Girvin, M. E. Cage, and K. von Klitzing, *The Quantum Hall Effect*, Second Edition (Springer, New York, 1990).

[41] D. Yoshioka, *The Quantum Hall Effect* (Springer, Berlin, 2002).

[42] D. Bures, Trans. Am. Math. Soc. **135**, 199 (1969).

[43] *Geometric Phases in Physics*, edited by A. Shapere and F. Wilczek (World Scientific, Singapore, 1989).

[44] D. J. Thouless, *Topological Quantum Numbers in Nonrelativistic Physics* (World Scientific, Singapore, 1998).

[45] A. Bohm, A. Mostafazadeh, H. Koizumi, Q. Niu, and J. Zwanzinger, *The Geometric Phase in Quantum Systems* (Springer, Berlin, 2003).

[46] J. J. Sakurai, *Modern Quantum Mechanics* (Addison-Wesley, Reading, 1994), p.140.

[47] D. J. Thouless, M. Kohmoto, M. P. Nightingale, and M. den Nijs, Phys. Rev. Lett. **49**, 405 (1982).

[48] C. L. Kane and E. J. Mele, Phys. Rev. Lett. **95**, 226801 (2005).

[49] D. N. Sheng, Z. Y. Weng, L. Sheng, and F. D. M. Haldane, Phys. Rev. Lett. **97**, 036808 (2006).

[50] S.-C. Zhang, Physics **1**, 6 (2008).

[51] Y. L. Chen et al., Science **325**, 178 (2009).

[52] J. E. Moore, Physics **2**, 82 (2009).

[53] X.-L. Qi and S.-C. Zhang, Phys. Today **63**(1), 38 (2010).

[54] J. P. Provost and G. Vallee, Commun. Math Phys. **76**, 289 (1980).

[55] F. Wilczek and A. Zee, Phys. Rev. Lett. **52**, 2111 (1984); reprinted in:*Geometric Phases in Physics*, edited by A. Shapere and F. Wilczek (World Scientific, Singapore, 1989), p. 141.

[56] `http://www.abinit.org/`.

[57] `http://www.crystal.unito.it/`.

[58] `http://cms.mpi.univie.ac.at/vasp/`.

[59] `http://www.quantum-espresso.org`.

[60] M.-C. Chang and Q. Niu, Phys. Rev. B **53**, 7010 (1996).

[61] G. Sundaram and Q. Niu, Phys. Rev. B **59**, 14915 (1999).

[62] C. A. Mead, Rev. Mod. Phys. **64**, 51 (1992).

[63] M. Kohmoto, Ann. Phys. **160**, 343 (1985).

[64] D. N. Zubarev, *Non-Equilibrium Statistical Mechanics* (Consultants Bureau, New York, 1974).

[65] P. Schmelcher, L. S. Cederbaum, and H.–D. Meyer, Phys. Rev. A **38**, 6066 (1988).

[66] L. Yin and C. A. Mead, J. Chem. Phys. **100**, 8125 (1994).

[67] D. Ceresoli, R. Marchetti and E. Tosatti, Phys. Rev. B **75**, 161101(R) (2007).

[68] R. Karplus and J. M. Luttinger, Phys. Rev. **95**, 1154 (1954).

[69] T. Jungwirth, Q. Niu, and A. H MacDonald, Phys. Rev. Lett. **88**, 207208 (2002).

[70] M. Onoda and N. Nagaosa, J. Phys. Soc. Jpn. **71**, 19 (2002).

[71] Y. Yao, L. Kleinman, A. H. MacDonald, J. Sinova, T. Jungwirth, D.-S. Wang, E. Wang, and Q. Niu, Phys. Rev. Lett. **92**, 037204 (2004).

[72] X. Wang, J. R. Yates, I. Souza, and D. Vanderbilt, Phys. Rev. B **74**, 195118 (2006).

[73] F. D. M. Haldane, Phys. Rev. Lett. **93**, 206602 (2004).

[74] X. Wang, D. Vanderbilt, J. R. Yates, and I. Souza, Phys. Rev. B **76**, 195109 (2007).

[75] N. W. Ashcroft and N. D. Mermin, *Solid State Physics* (Saunders, Philadelphia, 1976).

[76] J. C. Slater, Phys. Rev. **76**, 1592 (1959).

[77] J. M. Luttinger, Phys. Rev. **84**, 814 (1951).

[78] J. Zak, Phys. Rev. **168**, 686 (1968).

[79] D. Xiao, J. Shi, and Q. Niu, Phys. Rev. Lett. **95**, 137204 (2005).

[80] P. Streda, J. Phys. C **15**, L717 (1982).

[81] D. J. Thouless, Phys. Rev. B **27**, 6083 (1983).

[82] J. B. Pendry and C. H. Hodges, J. Phys. C **17**, 1269 (1984).

[83] J. Meister and W. H. E. Schwarz, J. Phys. Chem. **98**, 8245 (1994).

[84] L. D. Landau and E. M. Lifshitz, *Electrodynamics of Continuous Media* (Pergamon Press, Oxford, 1984).

[85] J. D. Jackson, *Classical Electrodynamics* (Wiley, New York, 1975).

[86] C. Kittel, *Introduction to Solid State Physics*, 7th. edition (Wiley, New York, 1996).

[87] R. M. Martin, Phys. Rev. B **9**, 1998 (1974).

[88] M. Posternak, A. Baldereschi, A. Catellani and R. Resta, Phys. Rev. Lett. **64**, 1777 (1990).

[89] R. Resta, Ferroelectrics **136**, 51 (1992).

[90] R. D. King-Smith and D. Vanderbilt, Phys. Rev. B **47**, 1651 (1993).

[91] G. Ortíz and R. M. Martin, Phys. Rev. B **43**, 14202 (1994).

[92] R. Resta, Phys. Rev. Lett. **80**, 1800 (1998).

[93] `http://www.uam.es/departamentos/ciencias/fismateriac/siesta/`.

[94] `http://www.cpmd.org/`.

[95] R. Resta, Modelling Simul. Mater. Sci. Eng. **11**, R69 (2003).

[96] W. H. Duan and Z. R. Liu, Curr. Opin. Solid State Mater. Sci. **10**, 40 (2006).

[97] M. Rabe, and J.-M. Triscone, eds., *Physics of Ferroelectrics: a Modern Perspective*, Topics in Applied Physics Vol. **105**, Ch. H. Ahn, K. (Springer-Verlag, 2007).

[98] M.P. Marder, *Condensed Matter Physics* (Wiley, New York, 2000).

[99] I. Souza, J. Ìñìguez, and D. Vanderbilt, Phys. Rev. Lett. **89**, 117602 (2002).

[100] P. Umari and A. Pasquarello, Phys. Rev. Lett. **89**, 157602 (2002).

[101] R. Resta, M. Posternak, and A. Baldereschi, Phys. Rev. Lett. **70**, 1010 (1993).

[102] G. H. Wannier, Phys. Rev. **52**, 191 (1937).

[103] N. Marzari and D. Vanderbilt, Phys. Rev. B **56**, 12847 (1997).

[104] A. A. Mostofi, Y.-S. Lee, I. Souza, D. Vanderbilt, and N. Marzari, Comput. Phys. Commun. **178**, 685 (2008).

[105] J. R. Yates, C. J. Pickard, and F. Mauri, Phys. Rev. B **76**, 024401 (2007).

[106] R. Resta, J. Chem. Phys. **124**, 104104 (2006).

[107] E. I. Blount, in *Solid State Physics*, edited by H. Ehrenreich, F. Seitz and D. Turnbull, vol **13** (Academic, New York, 1962), p. 305.

[108] O. F. Mossotti, Memorie di Matematica e di Fisica della Società Italiana delle Scienze Residente in Modena, **24**, 49 (1850); R. Clausius, *Die Mechanische Behandlung der Electrica* (Vieweg, Berlin, 1879).

[109] A. Pasquarello and R. Resta, Phys. Rev. B **68**, 174302 (2003).

[110] I. Dabo, B. Kozinsky, N. E. Singh-Miller, and N. Marzari, Phys. Rev. B **77** , 115139 (2008).

[111] P. Umari, A. Dal Corso, and R. Resta, in: *Fundamental Physics of Ferroelectrics: 2001 Williamsburg Workshop*, H. Krakauer, ed. (AIP, Woodbury, New York, 2001), p. 107.

[112] Q. Niu, Phys. Rev. **33**, 5368 (1986).

[113] D. Vanderbilt and R. D. King-Smith, Phys. Rev. B **48**, 4442 (1993).

[114] K. N. Kudin, R. Car, and R. Resta, J. Chem. Phys. **127**, 194902 (2007).

[115] M. Stengel, Phys. Rev. B **84**, 205432 (2011).

[116] P. L. Silvestrelli, M. Bernasconi, and M. Parrinello, Chem. Phys. Lett. **277**, 478 (1997).

[117] A. Pasquarello and R. Car, Phys. Rev. Lett. **79**, 1766 (1997).

[118] X. Gonze, Ph. Ghosez, and R. W. Godby, Phys. Rev. Lett. **74**, 4035 (1995).

[119] S. Baroni, S. de Gironcoli, A. Dal Corso, and P. Giannozzi, Rev. Mod. Phys. **73**, 515 (2001).

[120] F. Bloch, Z. Phys. **52**, 555 (1928).

[121] A. H. Wilson, Proc. Roy. Soc. A **133**, 458, and **134**, 277 (1931).

[122] N. F. Mott, Proc. Phys. Soc. (London) **62**, 416 (1949).

[123] P. W. Anderson, Phys. Rev. **109**, 1492 (1958).

[124] N. Mott, *Metal-Insulator Transitions*, 2nd ed. (Taylor & Francis, London, 1990).

[125] E. Abrahams (Ed.), *50 Years of Anderson Localization*, (World Scientific, Singapore, 2010).

[126] W. Kohn, Phys. Rev. **133**, A171 (1964).

[127] W. Kohn, in *Many–Body Physics*, edited by C. DeWitt and R. Balian (Gordon and Breach, New York, 1968), p. 351.

[128] Only about 80 citations (ISI) in the 25 years from 1964 to 1991.

[129] R. Resta and S. Sorella, Phys. Rev. Lett. **82**, 370 (1999).

[130] I. Souza, T. Wilkens, and R. M. Martin, Phys. Rev. B **62**, 1666 (2000).

[131] R. Resta, Phys. Rev. Lett. **95**, 196805 (2005).

[132] D. N. Zubarev, Soviet Phys. Ushpekhi **3**, 320 (1960).

[133] R. McWeeny, *Methods of Molecular Quantum Mechanics*, Second Edition (Academic, London, 1992).

[134] R. Kubo, M. Toda, and N. Hashitsume, *Statistical Physics II, Nonequilibrium Statistical Mechanics*, Springer Series in Solid-State Sciences, Vol. **31**, (Springer, Berlin, 1985).

[135] D. Forster, *Hydrodynamic Fluctuations, Broken Symmetry, and Correlation Functions* (Benjamin, Reading, 1975).

[136] R. Resta, Phys. Rev. Lett. **96**, 137601 (2006).

[137] C. Sgiarovello, M. Peressi, and R. Resta, Phys. Rev. **64**, 115202 (2001).

[138] M. Veithen, X. Gonze, and Ph. Ghosez, Phys. Rev. B **66**, 235113 (2002).

[139] N. D. M. Hine and W. M. C. Foulkes, J. Phys: Condens. Matter **19**, 506212 (2007).

[140] A. Monari, G. L. Bendazzoli, and S. Evangelisti, J. Chem. Phys. **129**, 134104 (2008).

[141] C. Aebischer, D. Baeriswyl, and R. M. Noack , Phys. Rev. Lett. **86**, 468 (2001).

[142] V. Vetere, A. Monari, G.L. Bendazzoli, S. Evangelisti, and B. Paulus, J. Chem. Phys. **128**, 214701 (2008).

[143] L. Stella, C. Attaccalite, S. Sorella, and A. Rubio, Phys. Rev. B **84**, 245117 (2011).

[144] G. L. Bendazzoli, S. Evangelisti, A. Monari, and R. Resta, J. Chem. Phys. **133**, 064703 (2010).

[145] E. Akkermans, J. Math. Phys. **38**, 1781 (1997).

[146] W. Kohn, Phys. Rev. Lett. **2**, 393 (1959).

[147] J. des Cloizeaux, Phys. Rev. **135**, A685 (1964); *ibid.* **135**, A697 (1964).

[148] S. Ismail-Beigi and T.A. Arias, Phys. Rev. Lett. **82**, 2127 (1999).

[149] L. He and D. Vanderbilt, Phys. Rev. Lett. **86**, 5341 (2001).

[150] G. F. Giuliani and G. Vignale, *Quantum Theory of the Electron Liquid* (Cambridge University Press, Cambridge, 2005).

[151] W. Kohn, Phys. Rev. Lett. **76**, 3168 (1996).

[152] J. G. Àngyàn, Int. J. Quantum Chem. **109**, 2340 (2009).

[153] J. G. Àngyàn, Curr. Org. Chem. **15**, 3609 (2011).

[154] R. Resta and S. Sorella, Phys. Rev. Lett. **87**, 4738 (1995).

[155] T. Wilkens and R. M. Martin, Phys. Rev. B **63**, 235108 (2001).

[156] B. Kramer and A. MacKinnon, Rep. Prog. Phys. **56**, 1469 (1993).

[157] D. J. Thouless, Phys. Rep. **13**, 93 (1974).

[158] A. Lagendijk, B. van Tiggelen, and D. S. Wiersma, Phys. Today **62**(8), 24 (2009).